
IÉSEG
SCHOOL OF MANAGEMENT
Booking.com
Web Scraping
Presented by: Pooja Shah

Project Scope:

- Information collected by Web Scraping from **Booking.com** for the city of **Lille**
- Data extracted on monthly basis for a total of 12 months from **December 2020 to November 2021**
- Key variables include:
  - Hotel/Apartment Name
  - Location (in Lille)
  - Total Price in Euros (for the selected period i.e. month)
  - Number of nights
  - Number of persons
  - Room Type
  - Beds
  - Rating
  - Rating Title
  - Number of reviews
  - Distance from city centre
  - Booking start date
  - Booking end date
  - Is breakfast included in the price?
  - Is free cancellation option applicable?

IÉSEG SCHOOL OF MANAGEMENT

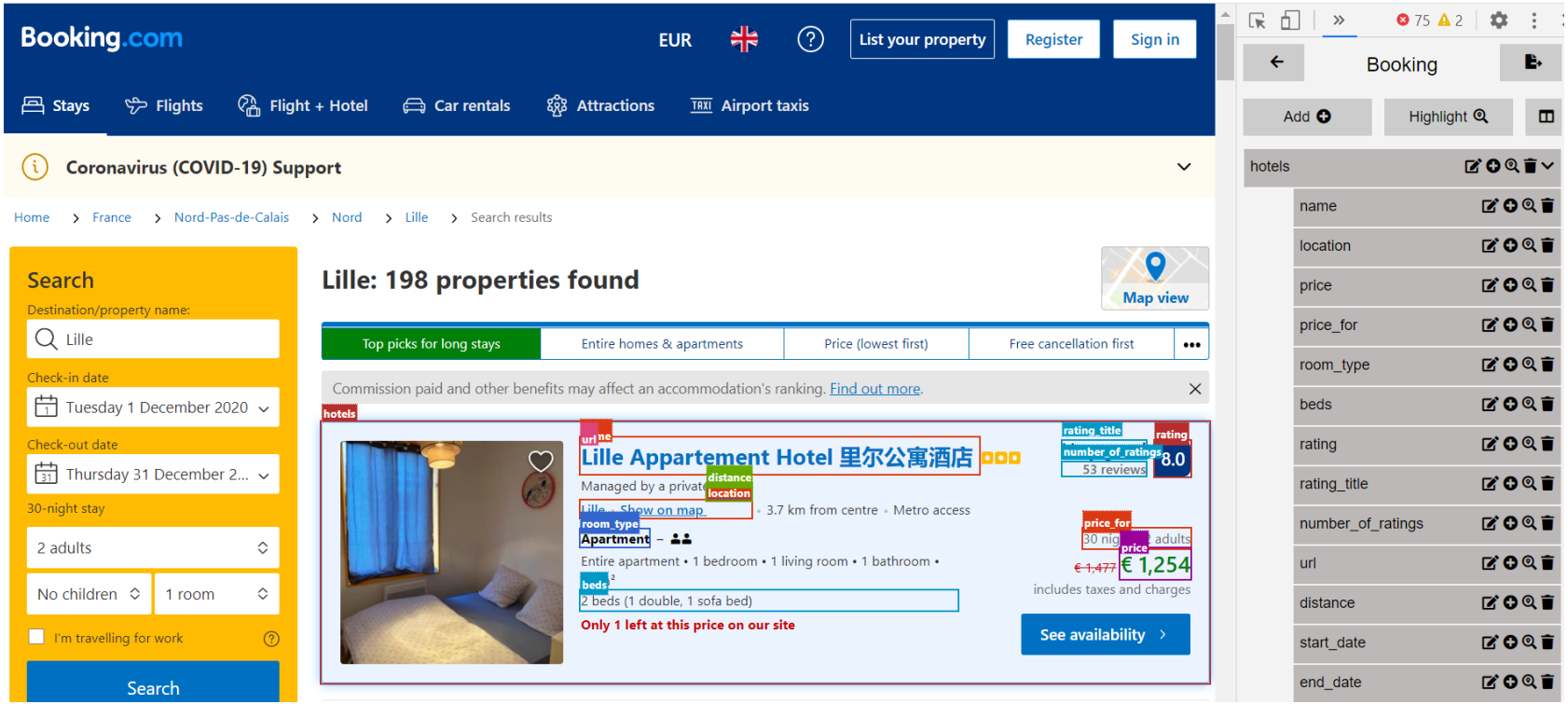
Project Flow:

- **Definition of Scope:** Based on the information available from Booking.com, I decided that I wanted to **focus** my analyses on 3 aspects for a duration of 1 year : Availability of accommodation options | Price | Ratings & Reviews
- **Definition of Goals:**
  - Have a clean usable data
  - Find relationships between several variables such as price, ratings, services offered, distance from city centre and more
- **Data collection :** After several trial and errors, I succeeded downloading a CSV using Selectorlib
- **Understanding, Cleaning & Processing:** Dropped some nulls for certain analysis, added new columns, converted data types, etc. using Pandas and Numpy mainly
- **Visualizations:** Plotted the graphs using Matplotlib, Seaborn and Plotly
- **Insights:** From the graphs and charts

IÉSEG SCHOOL OF MANAGEMENT

Challenges:

- Web scraping without any knowledge about it
- The biggest challenge was to fetch price data for hotels which was crucial for my analysis. With Beautiful Soup I could fetch all other data but not price. Hence I used Selectorlib to collect my final data
- Data cleaning and processing took time but was necessary to make data usable


Booking.com
EUR
List your property
Register
Sign in
Stays
Flights
Flight + Hotel
Car rentals
Attractions
Airport taxis
Coronavirus (COVID-19) Support
Home › France › Nord-Pas-de-Calais › Nord › Lille › Search results
Search
Destination/property name:
Lille
Check-in date
Tuesday 1 December 2020
Check-out date
Thursday 31 December 2...
30-night stay
2 adults
No children
1 room
I'm travelling for work
Search
Lille: 198 properties found
Map view
Top picks for long stays
Entire homes & apartments
Price (lowest first)
Free cancellation first
Commission paid and other benefits may affect an accommodation's ranking. Find out more.
hotels
Lille Appartement Hotel 里尔公寓酒店
Managed by a private
Lille – Show on map
3.7 km from centre
Metro access
Apartment
Entire apartment • 1 bedroom • 1 living room • 1 bathroom •
2 beds (1 double, 1 sofa bed)
Only 1 left at this price on our site
53 reviews
8.0
€ 1,477
€ 1,254
includes taxes and charges
See availability
Booking
Add
Highlight
hotels
name
location
price
price_for
room_type
beds
rating
rating_title
number_of_ratings
url
distance
start_date
end_date

Snapshot of Original Dataframe

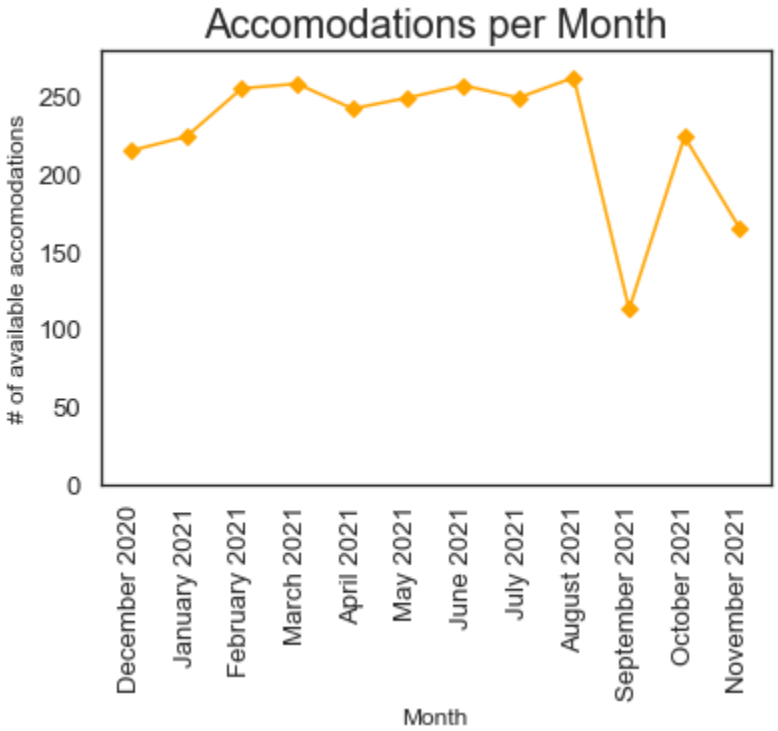
name	location	total_price	Nbr_of_nights	Nbr_of_persons	room_type	beds	rating	rating_title	number_of_ratings	distance_from_centre	start_date	end_date	breakfast_inclusion	free_cancellation	url
Deffrennes	Lille Sud	880	30	2	Budget Double Room	1 double bed	5.9	Review score	441 reviews	2.8 km	Tuesday 1 December 2020	Thursday 31 December 2020	Not Included	FREE cancellation	https://www.booking.com/hotel/fr/deffrennes.en...
Lille Appartement Hotel 里尔公寓酒店	Lille	1250	30	2	Apartment	2 beds(1 double, 1 sofa bed)	8.1	Very good	52 reviews	3.7 km	Tuesday 1 December 2020	Thursday 31 December 2020	Not Included	FREE cancellation	https://www.booking.com/hotel/fr/cozying-appar...
Little Suite - Manon	Vieux Lille	1260	30	2	Studio (2 Adults)	1 large double bed	NaN	NaN	NaN	550 m	Tuesday 1 December 2020	Thursday 31 December 2020	Not Included	FREE cancellation	https://www.booking.com/hotel/fr/little-suite-...
Nice apartment near EURALILLE	Lille Centre	1331	30	2	Studio	1 double bed	NaN	NaN	NaN	700 m	Tuesday 1 December 2020	Thursday 31 December 2020	Not Included	FREE cancellation	https://www.booking.com/hotel/fr/nice-apartmen...

2713 rows x **16** columns

IÉSEG SCHOOL OF MANAGEMENT


Pricing
Availability
Ratings

Room Availability



Insights:

There is a **sudden drop** in the availability of hotels in **September**. On closer inspection, it is observed that the average is drastically down only during the weekend of 4-5 September which turns out to be the "**Braderie de Lille**" weekend that attracts 3 million visitors every year.

The drop is thus explained by 2 possibilities:

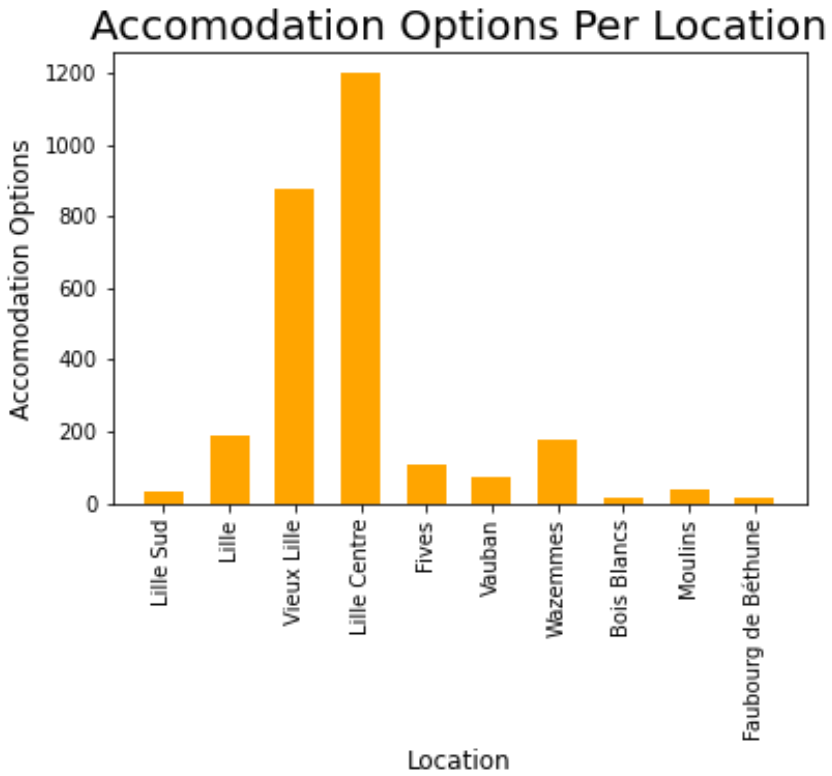
1. Many of the hotels have been already booked so much in advance
2. The hotels have chosen not to make the rooms available yet with hopes of higher prices in future (less likely)

IÉSEG SCHOOL OF MANAGEMENT

Hotels vs Apartments

- Availability:
  - Hotels < ~2/3rd Apartments
- September:
  - Either apartments are booked faster than hotels
  - Or apartment owners want to keep their apartment for themselves during "Braderie" weekend

Options by Areas



Insights:

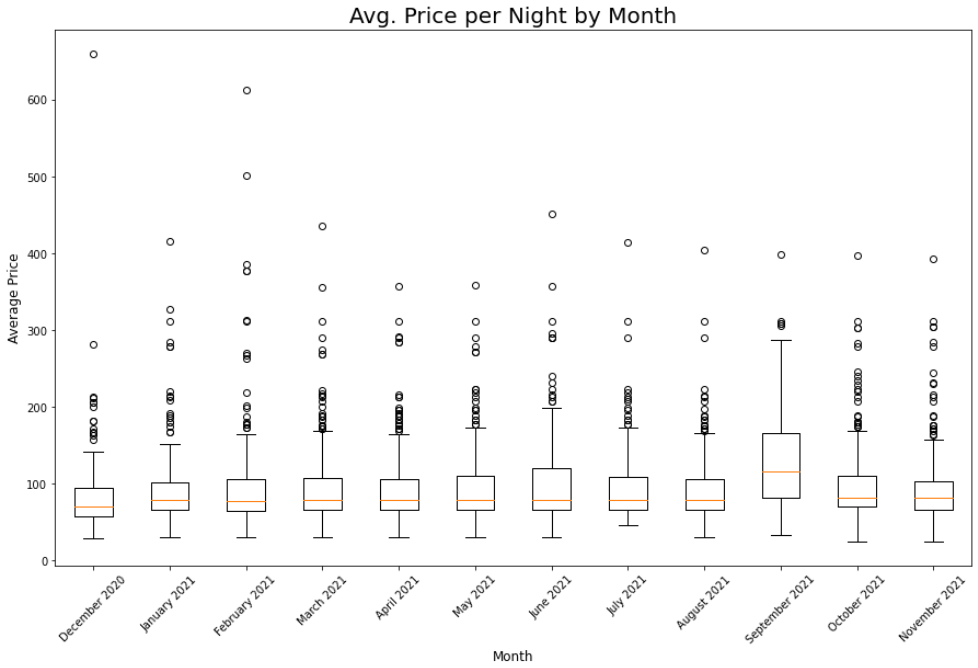
- As anticipated, Lille Centre has maximum number of hotels, followed by Vieux Lille in a duration of 12 months


Pricing
Availability
Ratings

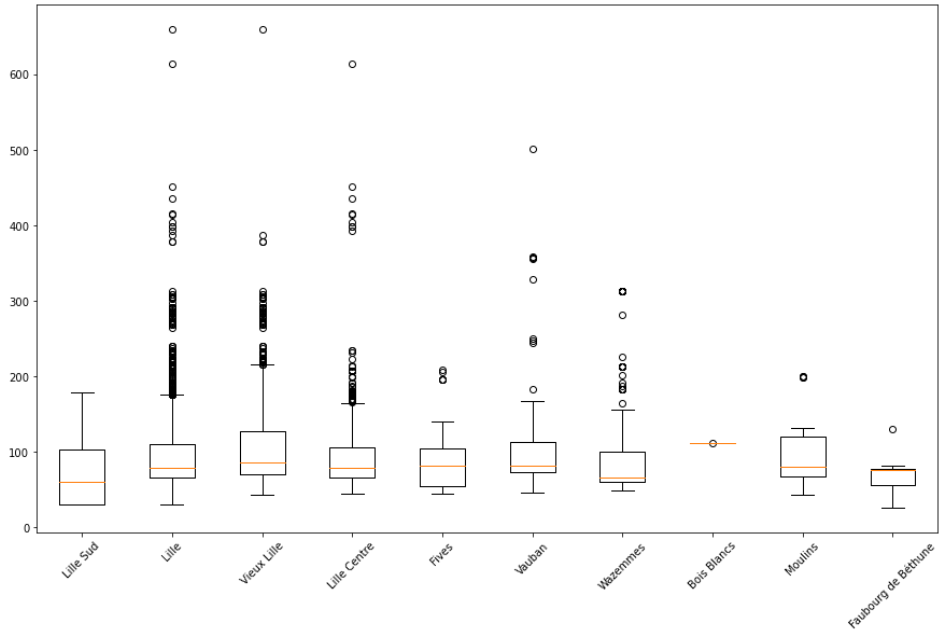
Price Per Night - Month

Insights:

- **Lower 50%** of the hotels have similar or **standard price range** throughout the year (except Sept.)
- The **top half** of the hotels have **varied prices** across all months
- September (being the festival month) and June (being the summer month) have a **larger third quartile** possibly because of **higher earning scope** during those months



Price Per Area



Insights:

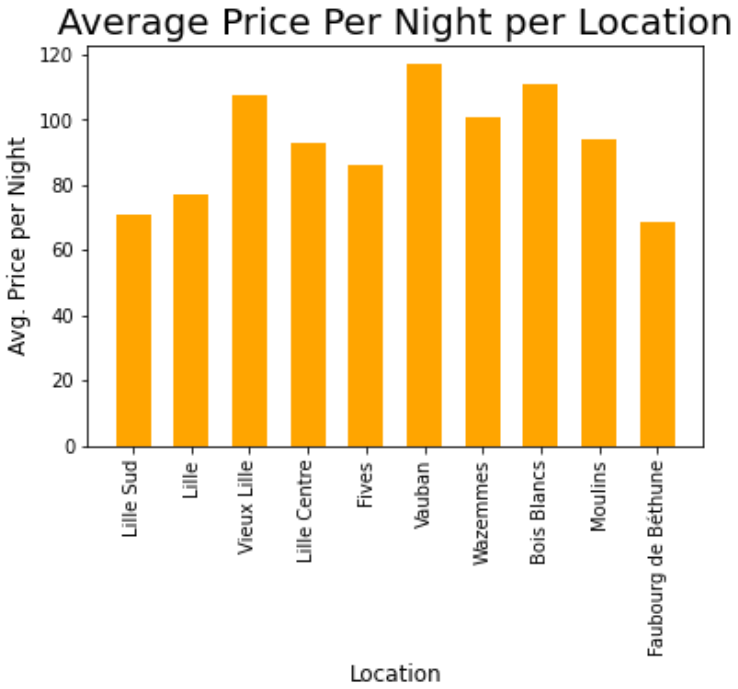
- **Bois Blancs** seems to have only 1 hotel which explains why its **median price is higher** than all other quartiers
- Top 50% of the hotels in **Vieux Lille** is higher priced compared to top 50% in other locations
- **Faubourg de Béthune** is a low-budget area

IÉSEG SCHOOL OF MANAGEMENT

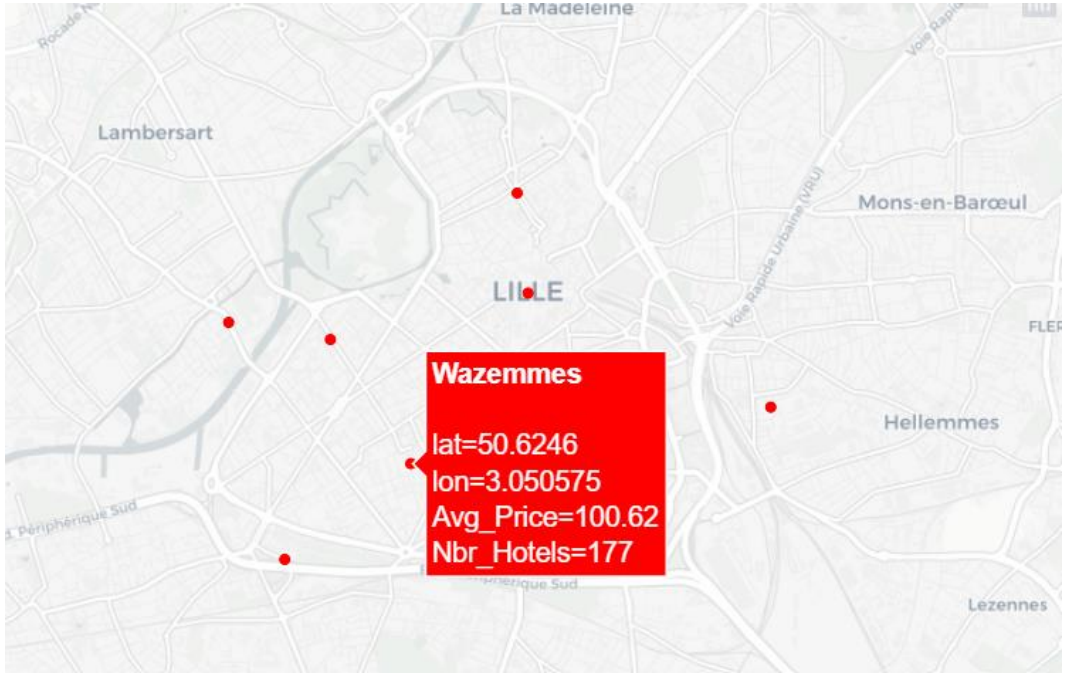
Price Per Area

Insights:

- As the previous boxplot suggests, **Faubourg de Béthune** has **lowest** hotel prices and **Bois Blancs** is **expensive** compared to most others because of a hotel's monopoly in the area
- However, while the boxplot gave an **impression** that possibly Lille Centre and Vieux Lille would be more expensive, it turns out that average price for **Vauban is higher** than all other areas

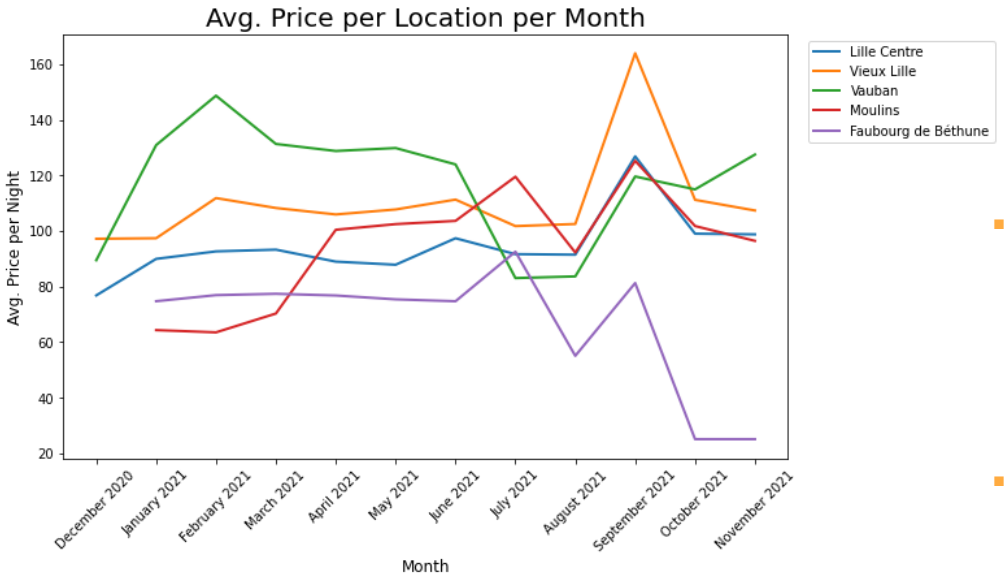


Price Per Area – Map



- Geolocation map with hover data with Plotly package

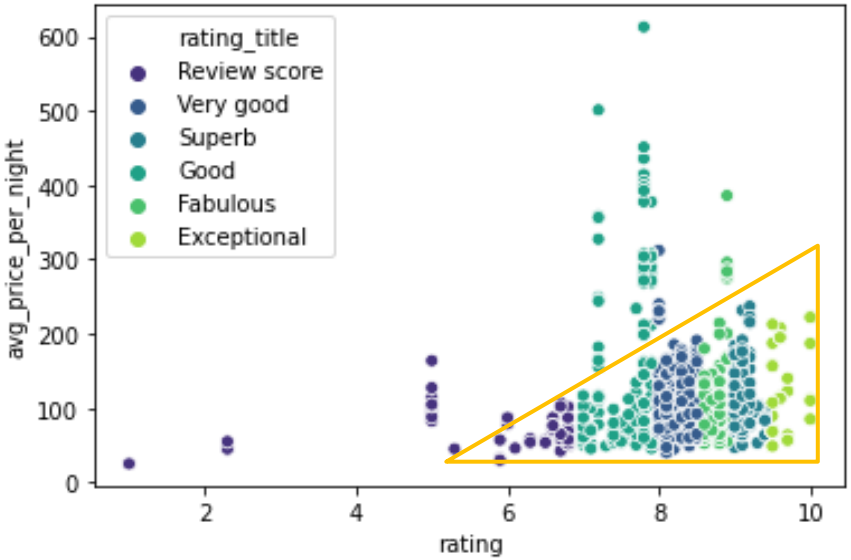
Price Per Area Per Month



- Vauban:
  - January- May, September onwards – High - student area with most universities & colleges
  - June-August – Low – Students go back home/on vacation during the summer break
- Vieux Lille and Lille Centre:
  - Similar trend as favourite tourist places
  - Peak in September is explained earlier - Braderie
- Faubourg de Béthune :
  - Low because of its reputation as among the poorest and unsafe areas


Pricing
Availability
Ratings

Ratings & Price



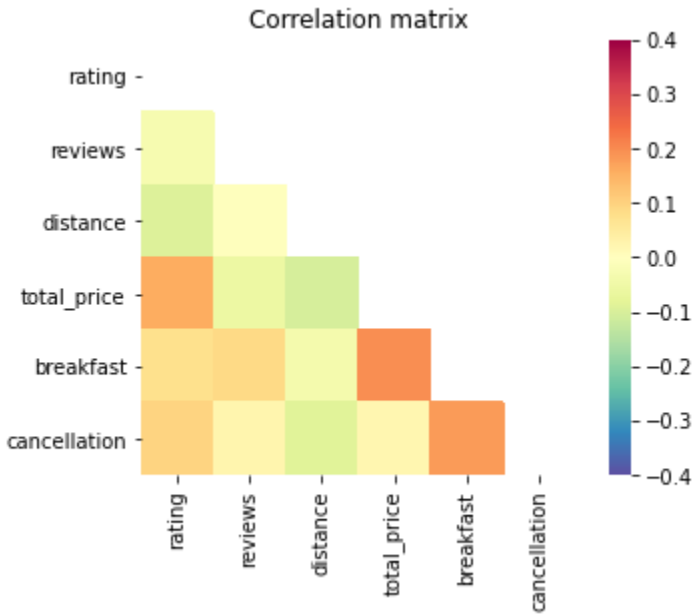
Insights:

- As the ratings increase there's also a corresponding change in price
- The "Very Good" hotels are super bunched up with the highest count of hotels and competitive prices
- Most of the high price outliers in the "Good" range are from 2-3 properties in Vieux Lille consistently priced higher than properties in the "Good" range in other locations

Rating	Count of Hotels
Review score	115
Very good	488
Superb	239
Good	209
Fabulous	233
Exceptional	61

IÉSEG SCHOOL OF MANAGEMENT

Correlation Matrix – Positive



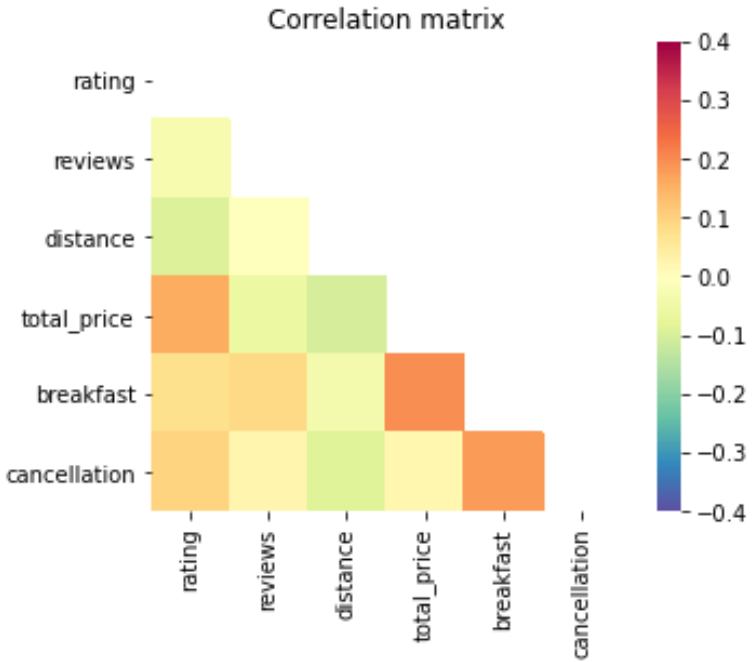
Insights:

- Direct correlation between inclusion of breakfast in the booking and the total price per night
- **Interesting:** Good correlation between 'breakfast' and 'cancellation', meaning that hotels that offer one of these services also tend to offer the other one
- Higher the ratings, higher the cost per night
- Slight positive correlation between free cancellation service and higher ratings

IÉSEG SCHOOL OF MANAGEMENT

Correlation Matrix – Negative

- As the hotels are closer to the city centre (distance), the ratings on an average are higher
- Hotels closer to the city centre are priced higher
- Closer the hotel to the centre, more options for free cancellation are available. Possible explanation: These hotels would have more assurance of being occupied even later while far away hotels cannot expect the same



Thank You!

Questions?