

Capstone Project 2

Data Analysis of Luxury Hotels Sector in Amsterdam

Contents:

- Problem Statement
- ER Diagram
- Relational Schema
- Data Collection
- Data Preparation
- Creating tables & Loading data into SQL
- Data Analysis using SQL
- Assumptions & Limitations
- Locations of the Hotels in Amsterdam
- Dashboard
- Conclusion



Problem Statement:

The Stakeholders :

- The Netherlands Tourism Board
- The City Government of Amsterdam

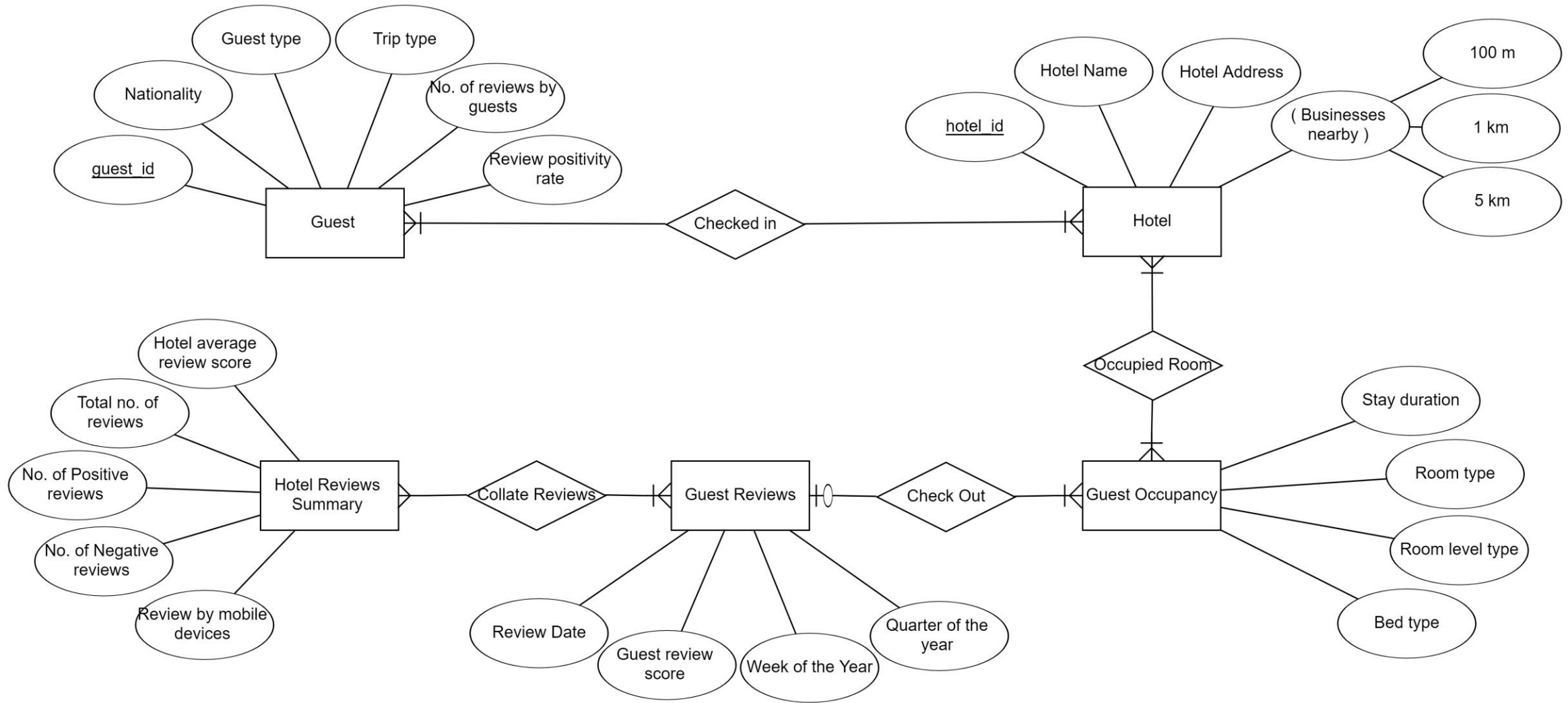


Problem Definition :

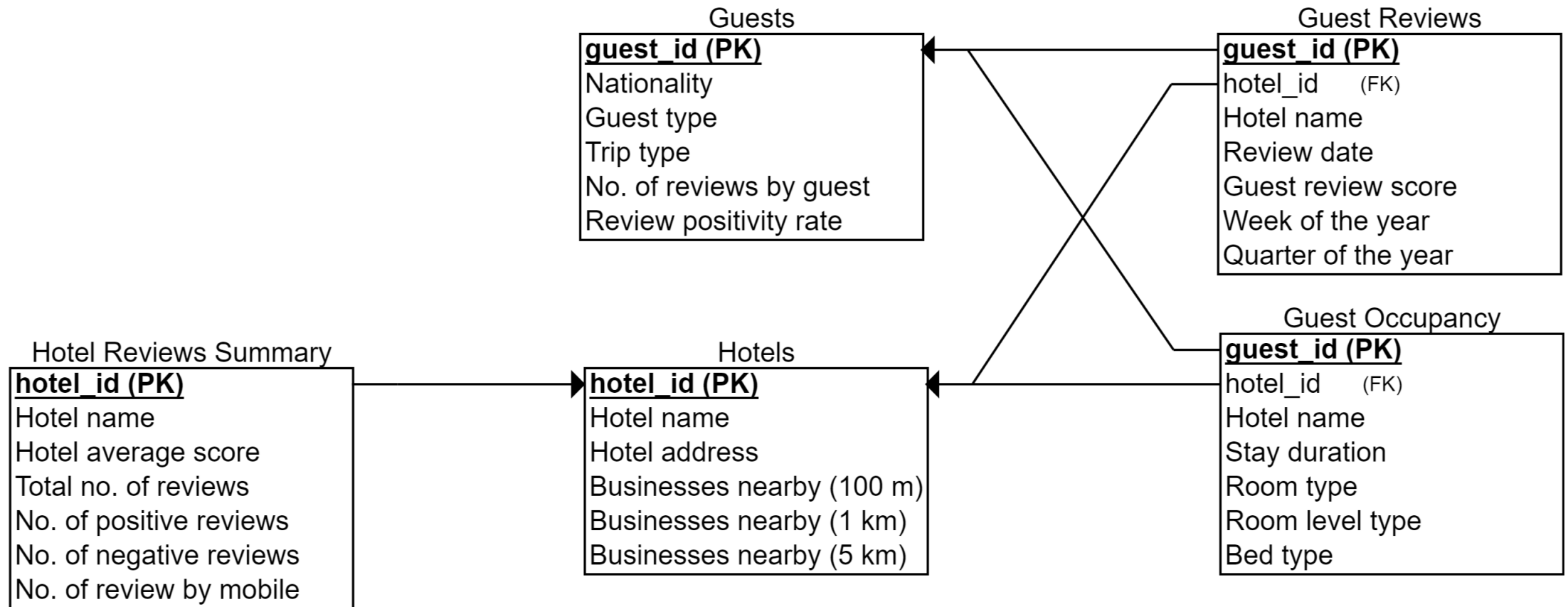
- Who comes to Amsterdam?
- Purpose of travel to Amsterdam?
- Any potential sector to develop and diversify?
- How to continue the growth of the Luxury Hotel Sector?



ER Diagram:

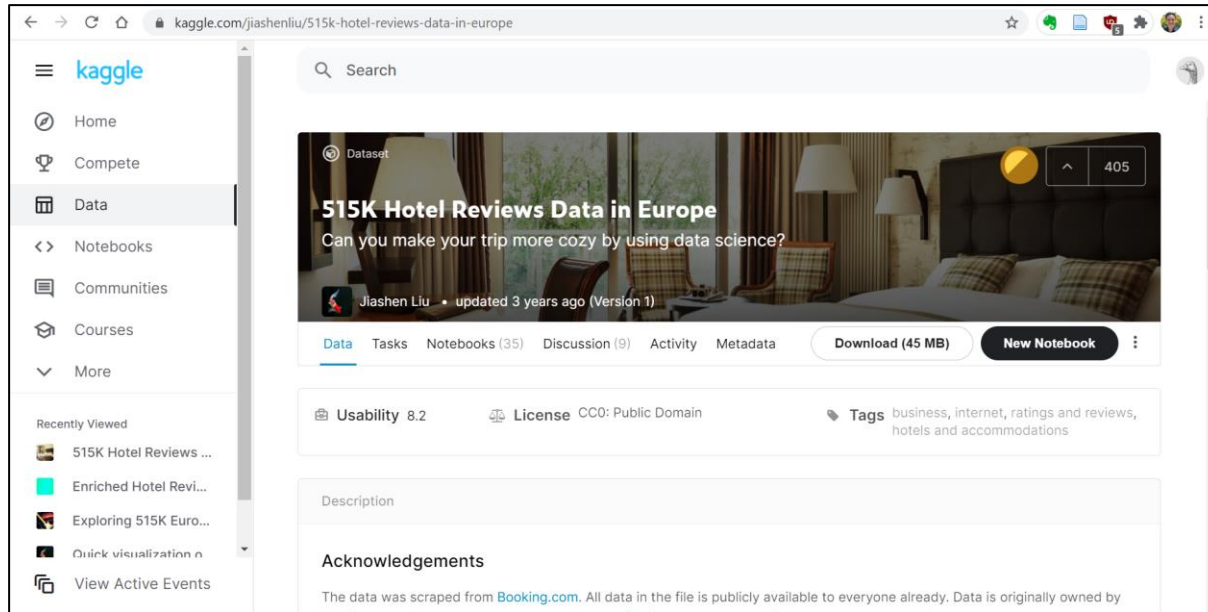


Relational Schema:

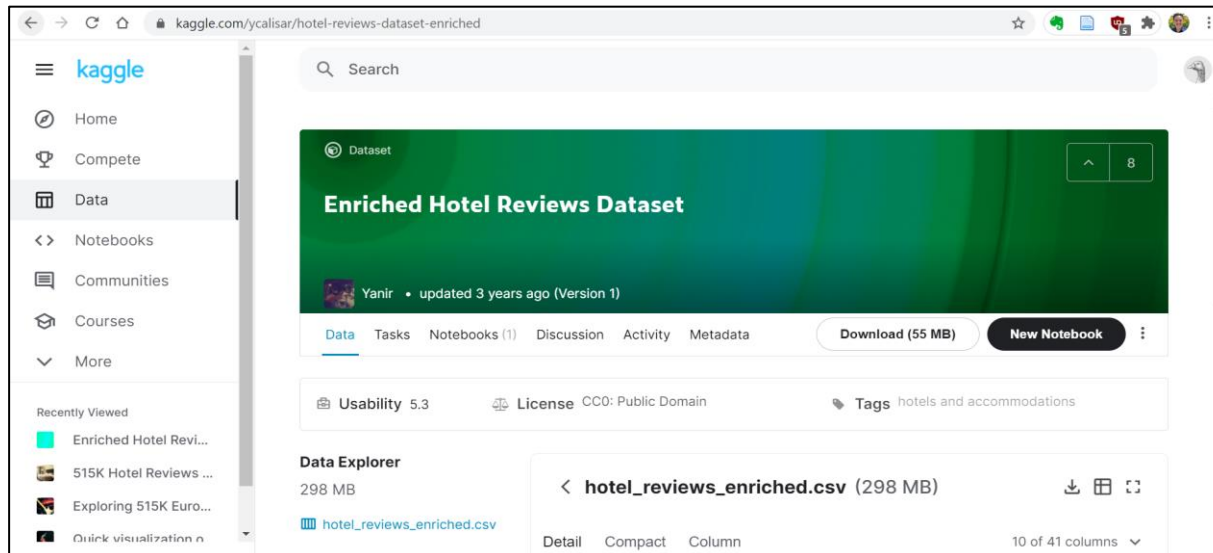


Data collection:

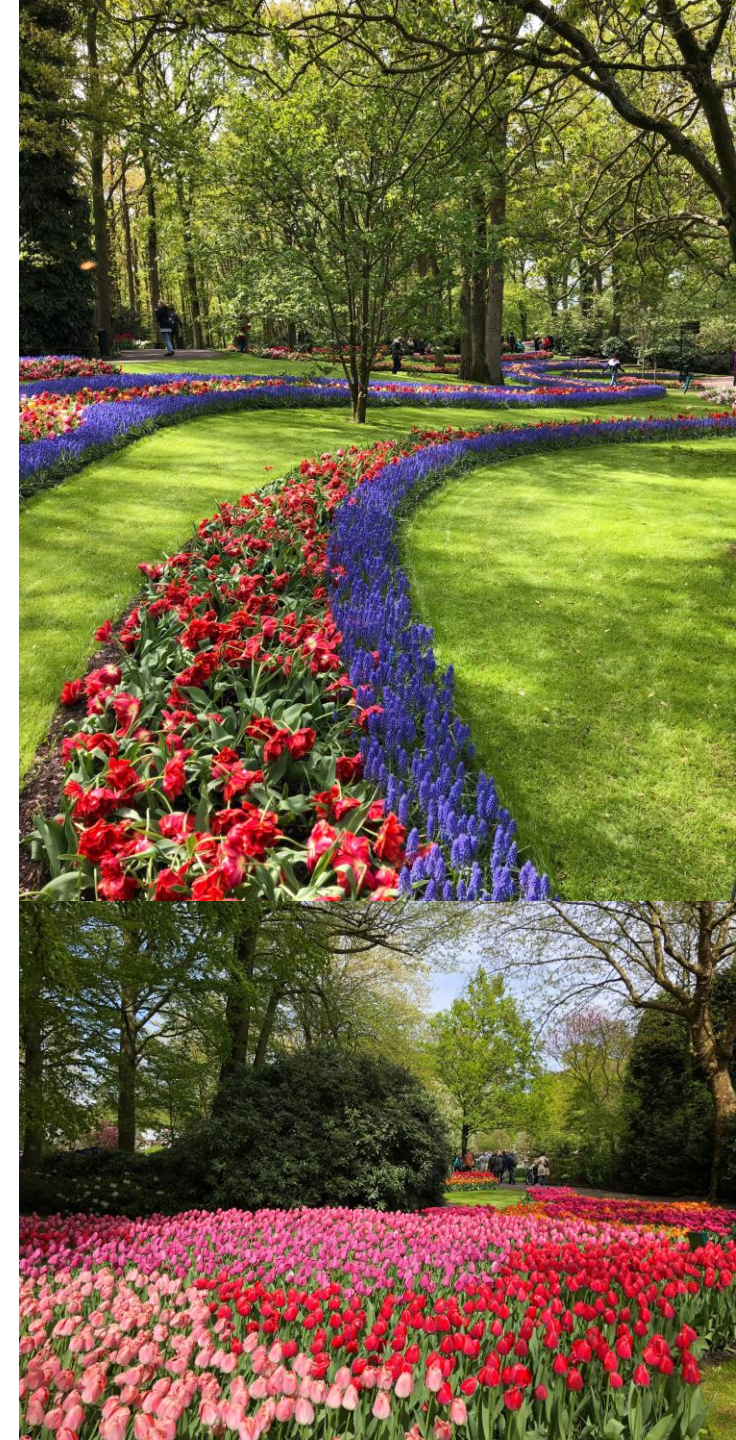
- Data Collection from Kaggle website



This screenshot shows the Kaggle dataset page for '515K Hotel Reviews Data in Europe' by Jiashen Liu. The page features a header with the Kaggle logo and navigation links. The dataset title is prominently displayed, along with a description: 'Can you make your trip more cozy by using data science?'. Below the title, there are tabs for 'Data', 'Tasks', 'Notebooks (35)', 'Discussion (9)', 'Activity', and 'Metadata'. The 'Data' tab is selected, showing a 'Download (45 MB)' button and a 'New Notebook' button. The dataset's usability is rated 8.2, and it is licensed under CC0: Public Domain. The tags include 'business', 'internet', 'ratings and reviews', 'hotels and accommodations'. The description section is partially visible, mentioning that the data was scraped from Booking.com.



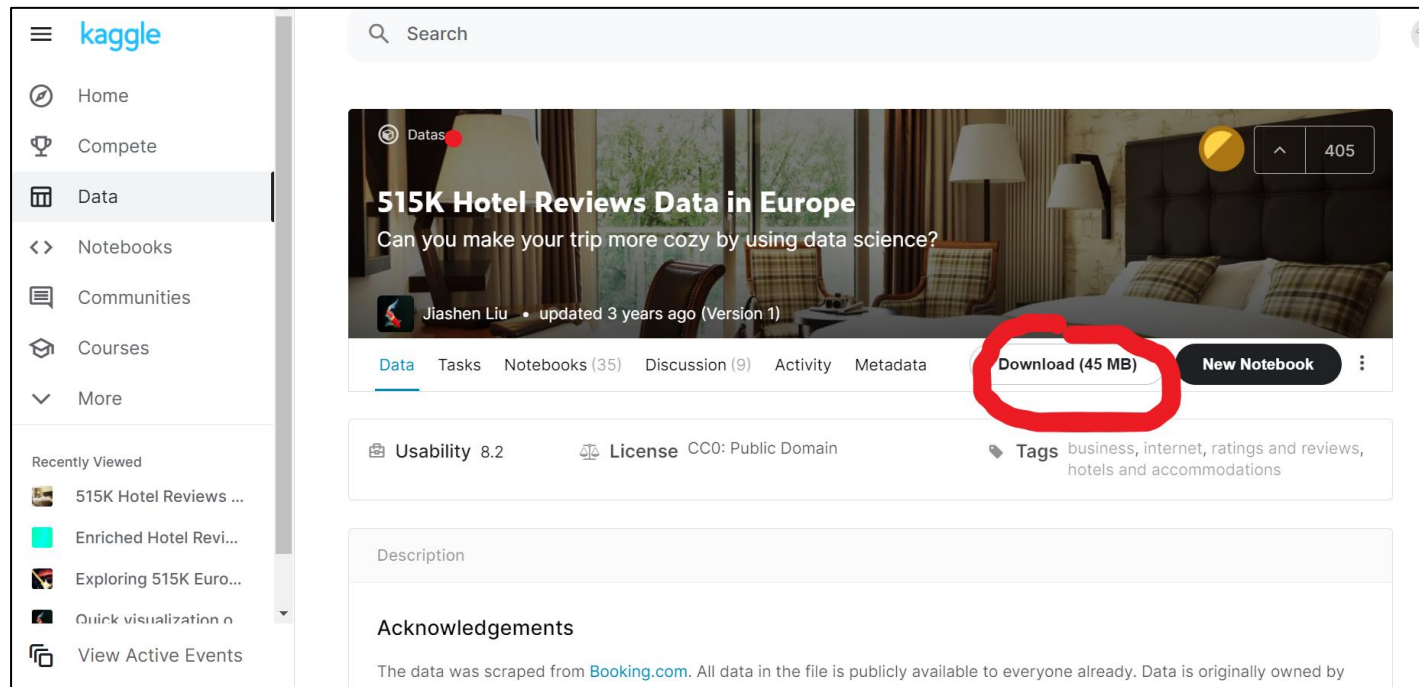
This screenshot shows the Kaggle dataset page for 'Enriched Hotel Reviews Dataset' by Yanir. The page features a header with the Kaggle logo and navigation links. The dataset title is prominently displayed, along with a description: 'Can you make your trip more cozy by using data science?'. Below the title, there are tabs for 'Data', 'Tasks', 'Notebooks (1)', 'Discussion', 'Activity', and 'Metadata'. The 'Data' tab is selected, showing a 'Download (55 MB)' button and a 'New Notebook' button. The dataset's usability is rated 5.3, and it is licensed under CC0: Public Domain. The tags include 'hotels and accommodations'. The description section is partially visible, mentioning that the data was scraped from Booking.com. Below the dataset information, there is a 'Data Explorer' section showing the file 'hotel_reviews_enriched.csv' (298 MB) with a download icon and a view icon. The file is listed as 'hotel_reviews_enriched.csv' and has 10 of 41 columns.



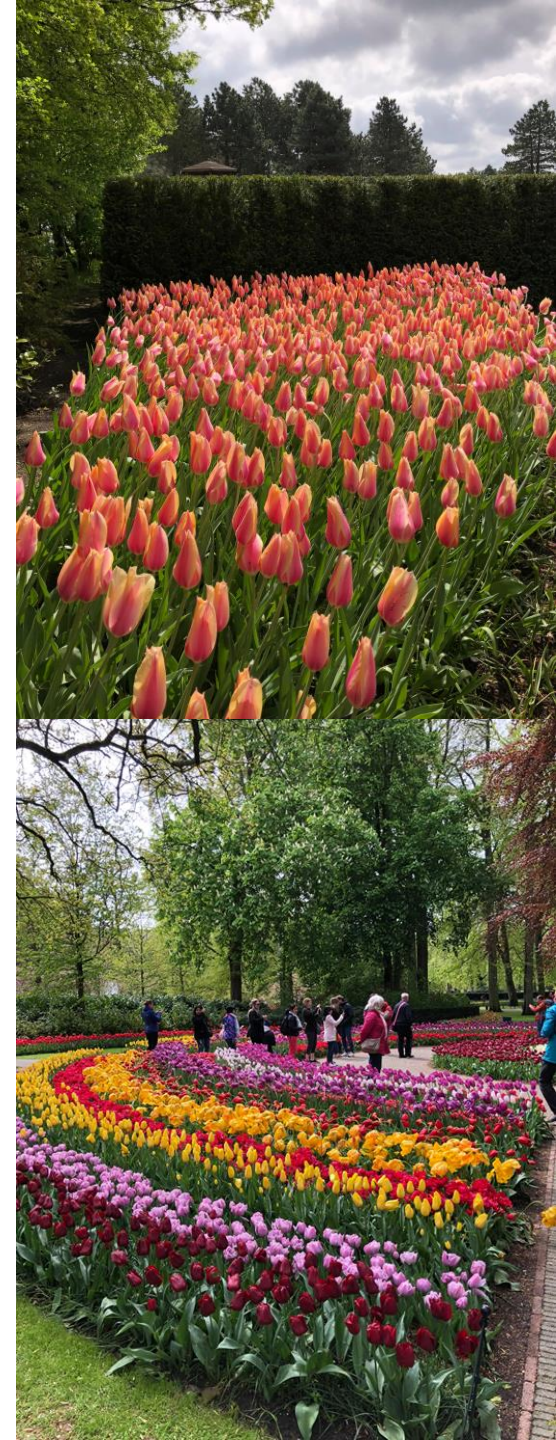
Data Preparation:

- Data Transformation:

1. Download data file from Kaggle

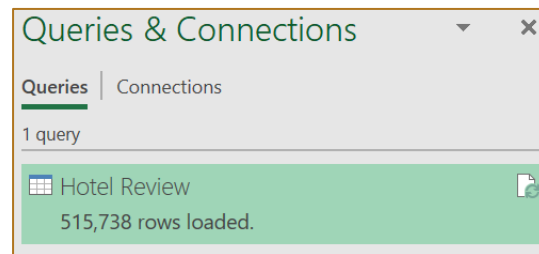
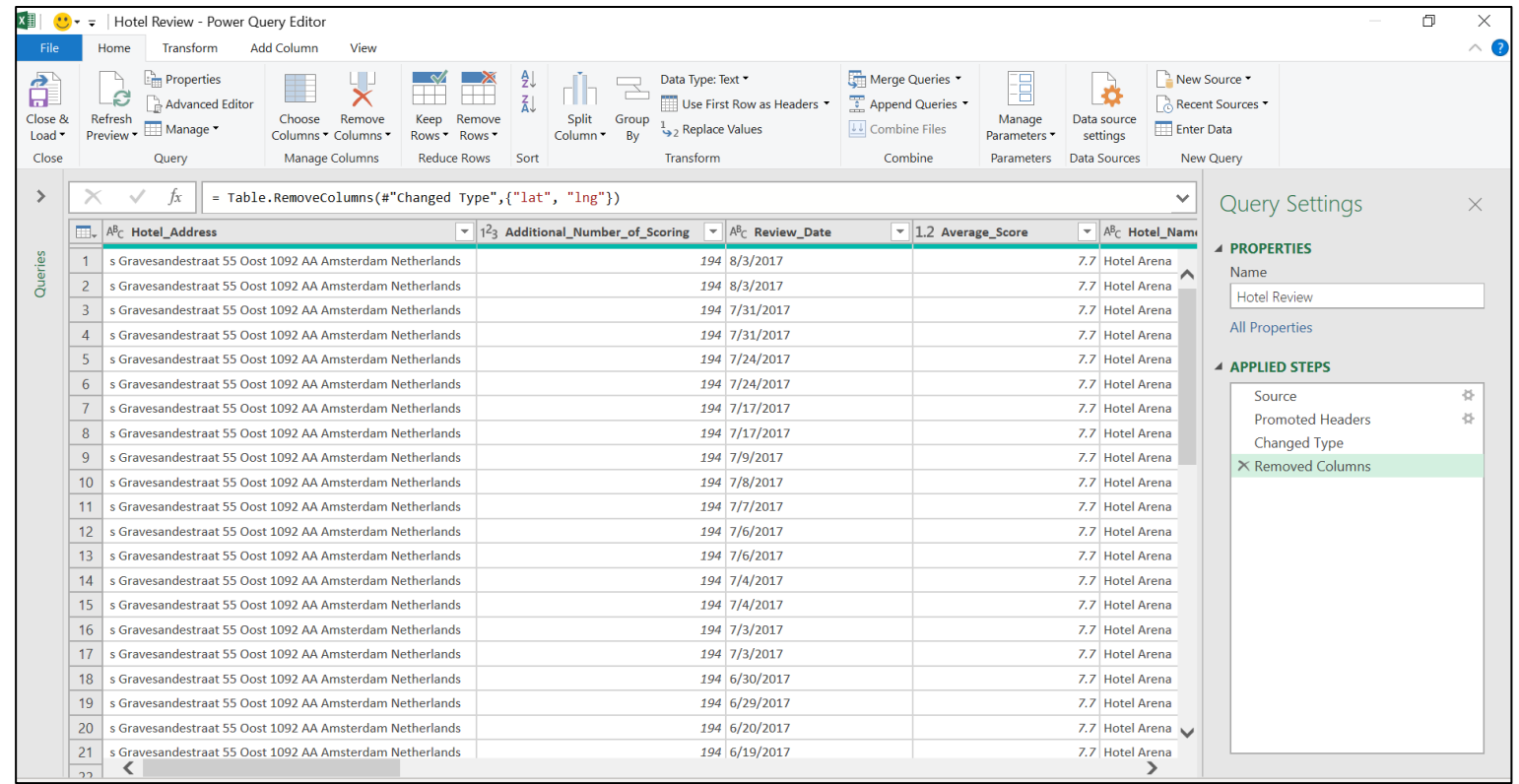
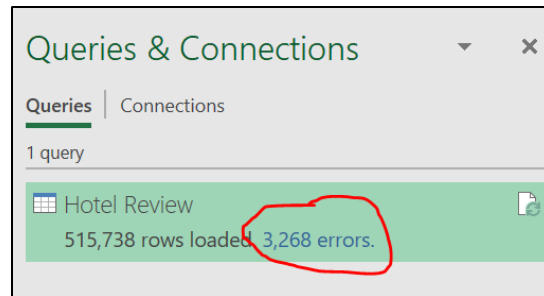


2. Changed file name



Data Preparation:

3. Data transformation through power query (e.g. Removed errors by removing problem Columns etc.)



Data Preparation:

4. Uploaded it to SQL Server

Import Flat File 'CapstoneProject2'

Modify Columns

Introduction
Specify Input File
Preview Data
Modify Columns
Summary
Results

Modify Columns
This operation generated the following table schema. Please verify if schema is accurate, and if not, please make any changes.

Column Name	Data Type	Primary Key	Allow Nulls
Hotel_Address	nvarchar(100)	<input type="checkbox"/>	<input type="checkbox"/>
Additional_Number_of_Scoring	int	<input type="checkbox"/>	<input type="checkbox"/>
Review_Date	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Average_Score	float	<input type="checkbox"/>	<input type="checkbox"/>
Hotel_Name	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Reviewer_Nationality	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Negative_Review	nvarchar(MAX)	<input type="checkbox"/>	<input type="checkbox"/>
Review_Total_Negative_Word_Counts	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Total_Number_of_Reviews	int	<input type="checkbox"/>	<input type="checkbox"/>
Positive_Review	nvarchar(MAX)	<input type="checkbox"/>	<input type="checkbox"/>
Review_Total_Positive_Word_Counts	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Total_Number_of_Reviews_Reviewer_Has_Given	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>
Reviewer_Score	float	<input type="checkbox"/>	<input type="checkbox"/>
Tags	nvarchar(MAX)	<input type="checkbox"/>	<input type="checkbox"/>
days_since_review	nvarchar(50)	<input type="checkbox"/>	<input type="checkbox"/>

Error Reporting - Selecting a smaller range may have a performance impact

No Range

< Previous Next > Cancel

- Data Cleaning

```
SELECT * FROM HotelReview  
  
ALTER TABLE HotelReview  
ALTER COLUMN Review_Date VARCHAR(20)  
  
ALTER TABLE HotelReview  
ALTER COLUMN Review_Date DATE
```

```
SELECT DISTINCT Average_score,  
SUM (Sum_of_Count_Review_Is_Negative) AS 'Total no. of negative reviews'  
FROM Hotels  
JOIN Hotel_Review_Summary  
ON Hotels.hotel_id = Hotel_Review_Summary.hotel_id  
GROUP BY Average_score  
ORDER BY Average_score;
```

Creating tables & Loading data into SQL:

```
CREATE TABLE
Guests
(guest_id int PRIMARY KEY,
Nationality varchar(50),
Guest_type varchar(50),
Trip_type varchar(50),
No_of_reviews_by_guests int,
Review_positivity_rate varchar(10) );

INSERT INTO Guests (guest_id, Nationality, Guest_type, Trip_type, No_of_reviews_by_guests, Review_positivity_rate)
SELECT guest_id, Reviewer_Nationality, Guest_Type, Trip_Type, Total_Number_of_Reviews_Reviewer_Has_Given, Review_Positivity_Rate
FROM HotelReview;
```

```
CREATE TABLE
Hotels
(hotel_id INT identity(1,1) PRIMARY KEY,
Hotel_name VARCHAR(50),
Hotel_address VARCHAR(MAX),
Businesses_nearby_100m INT,
Businesses_nearby_1km INT,
Businesses_nearby_5km INT);

INSERT INTO Hotels(Hotel_name, Hotel_address, Businesses_nearby_100m, Businesses_nearby_1km, Businesses_nearby_5km)
SELECT DISTINCT Hotel_Name, Hotel_Address, Businesses_100m, Businesses_1km, Businesses_5km
FROM HotelReview;
```


Data Analysis using SQL:

```
= SELECT TOP 10 Nationality,  
COUNT(*) AS 'Guest Numbers'  
FROM Guests  
GROUP BY Nationality  
ORDER BY 'Guest Numbers' DESC;
```

```
= SELECT Nationality,  
ROUND(AVG(Review_positivity_rate),2) AS 'Positivity_rate_by_nationality'  
FROM Guests  
WHERE Nationality  
LIKE 'United%'  
GROUP BY Nationality  
ORDER BY 'Positivity_rate_by_nationality' DESC;
```

```
= SELECT DISTINCT Average_score,  
SUM(Sum_of_Count_Review_Is_Positive) AS 'Total no. of positive reviews'  
FROM Hotels  
JOIN Hotel_Review_Summary  
ON Hotels.hotel_id = Hotel_Review_Summary.hotel_id  
GROUP BY Average_score  
ORDER BY Average_score;
```

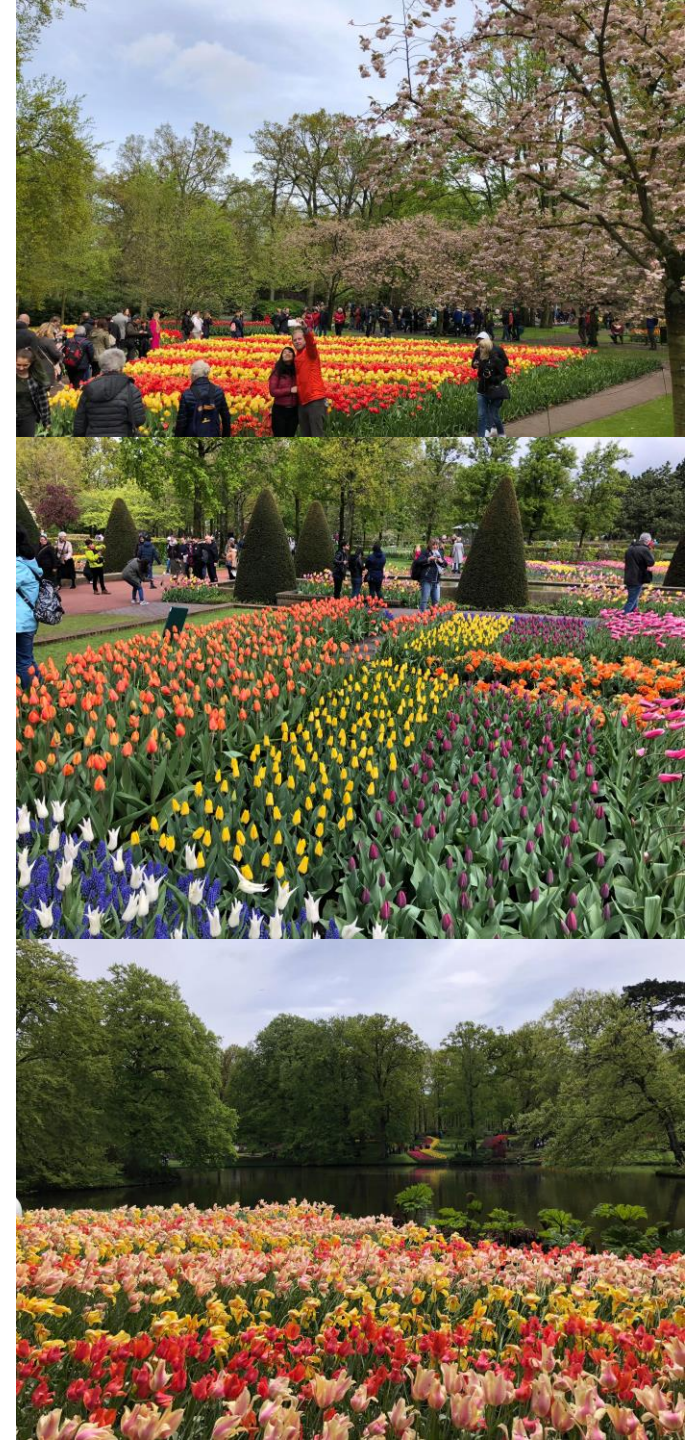
Assumptions & Limitations:

Assumptions :

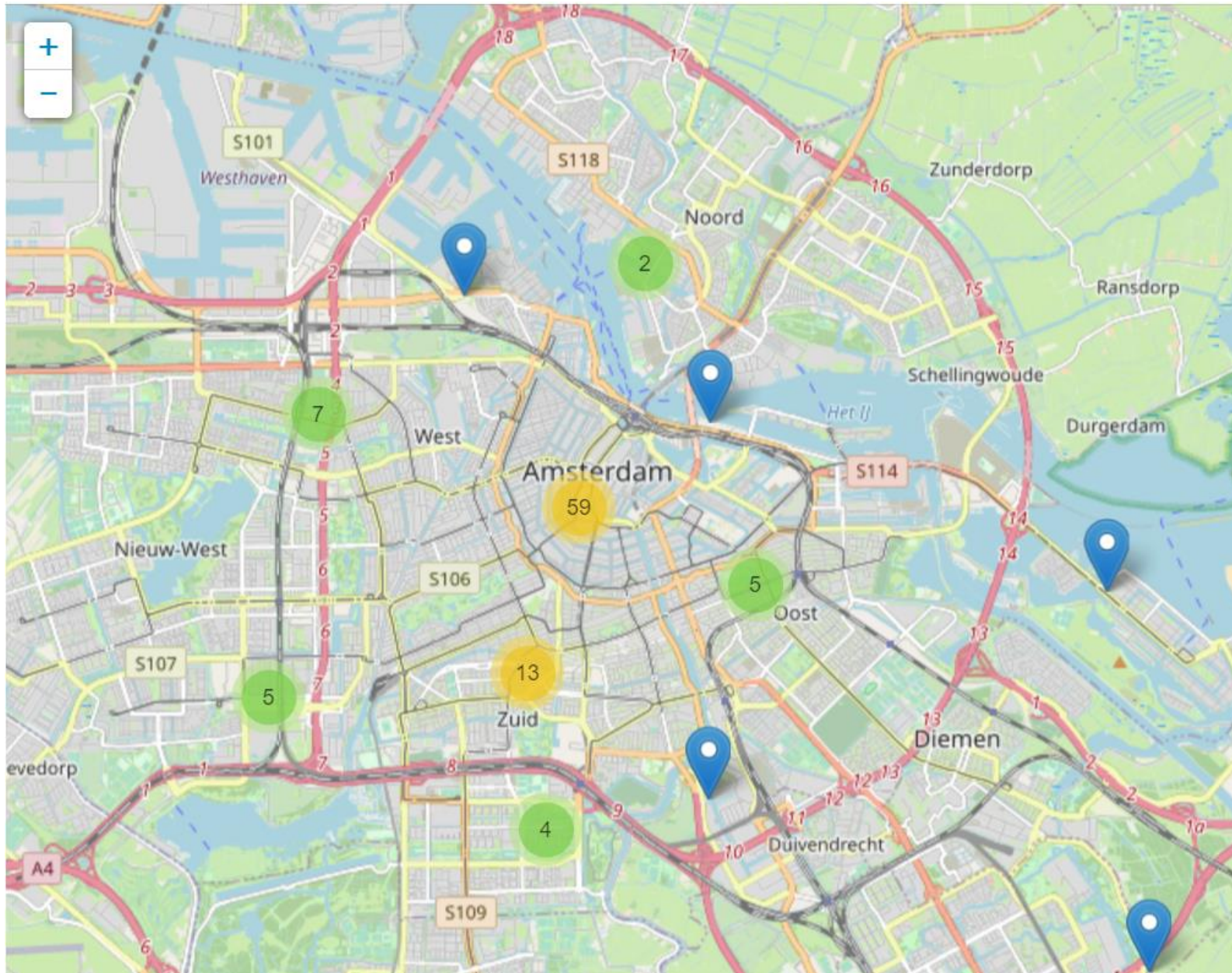
- No repeat guest reviews
- No bias reviews

Limitations :

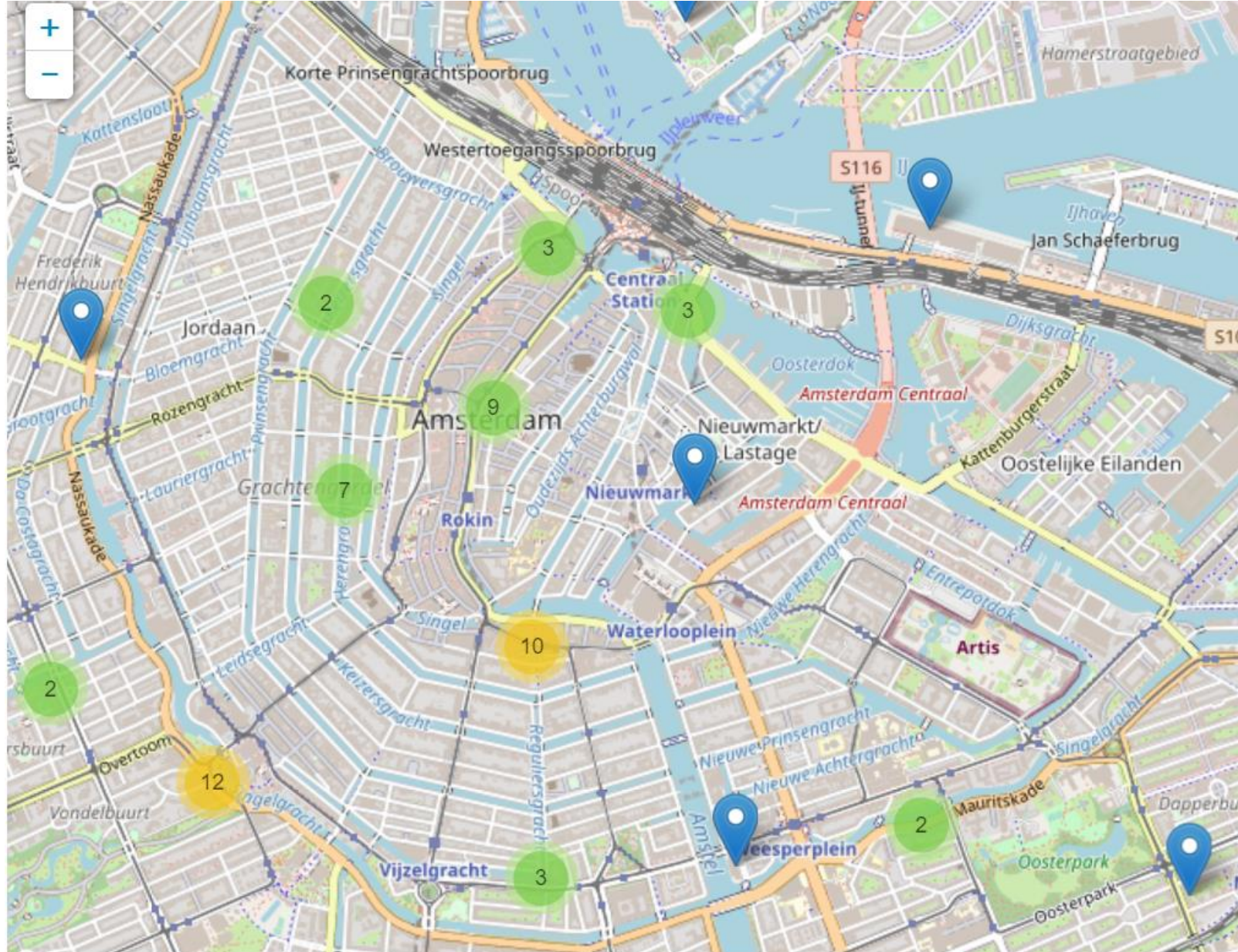
- Only between year 2015 to 2017
- Only from one source of dataset



Locations of the Hotels in Amsterdam:



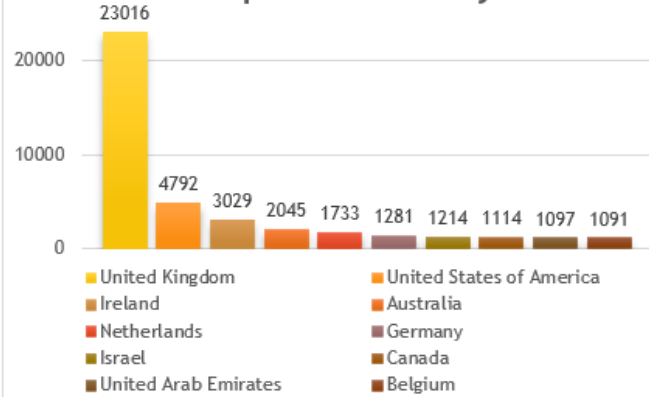
Locations of the Hotels in Amsterdam:



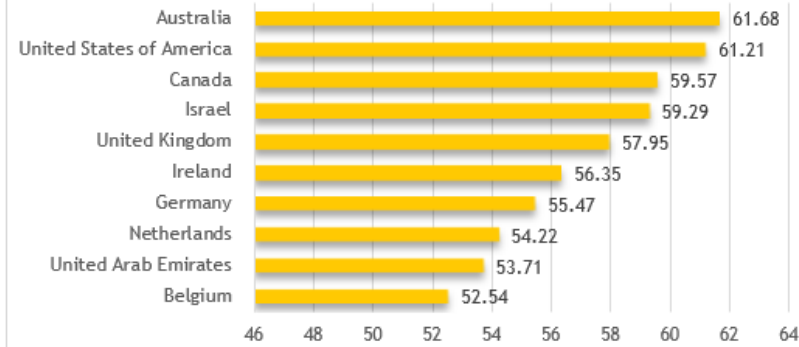
Dashboard:

Dashboard Analysis of Luxury Hotel Sector in Amsterdam from 2015 to 2017

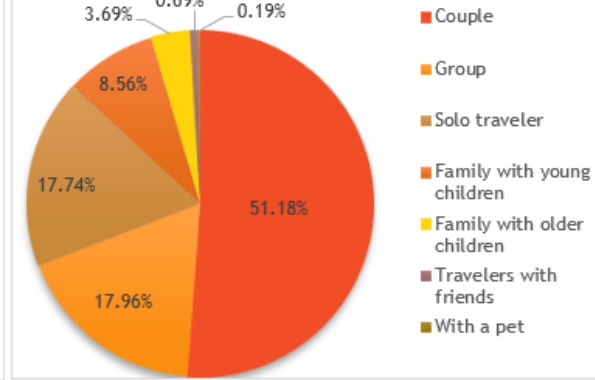
Top 10 Nationality



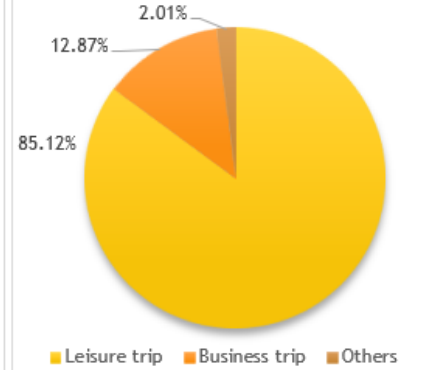
Top 10 Nationality Review Postivity Rate



Guest Type



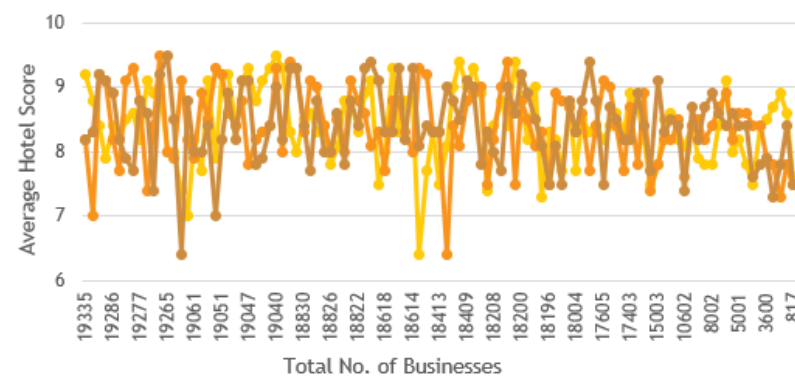
Trip Type



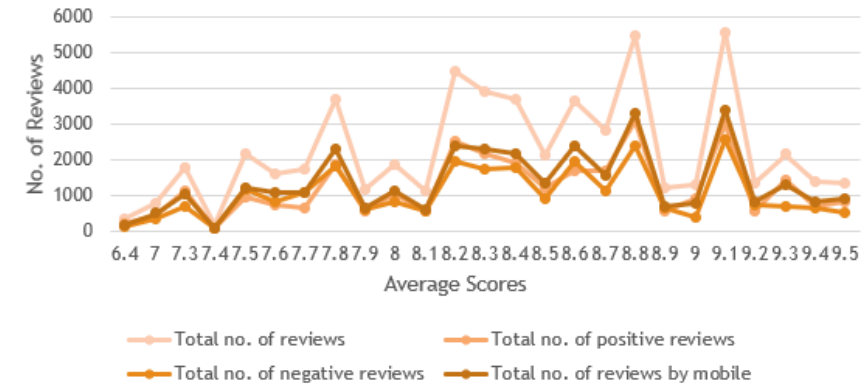
No. of check-in per week



Number of Businesses Nearby



Hotel Average Score



No. of Businesses

100m (1 - 135 Busi... 1km (9 - 4000 Busi... 5km (408 - 9600 Bu...

Average_score

Total no. of negative reviews Total no. of positive reviews
Total no. of reviews Total no. of reviews by mobile

Conclusion:

- **Insights:**

- Large Proportion of guests from the UK & English speaking countries
- More than 50% of guest types are couples
- Huge percentage are Leisure trip guests at 85.12%
- Trend of higher no. of reviews that leads to higher hotel scores

- **Potential areas for development:**

- Increase guest numbers for Autumn seasons
(E.g. Emphasise on lower crowd levels)
- Encourage more solo and group travellers
(E.g. Special rates for Solo guest and groups)
- Improve hotel facility for business guests
(E.g. Strong & stable ultra high speed WiFi connections)
- Encourage guest to give hotel reviews
(E.g. Incentives to earn membership points or discounts)

