

# MiniProject2 - SQL: From Data to Insight

NYC Airbnb for STUDENTS  
**StudentHub: NYC Rentals**

Kinga, Julia, Joma, Priyanka

DATA-FT-Sept-2025

# Mini Project Overview:

1

Selecting NYC Dataset to analyse

- Data retrieved from Kaggle

2

Tools chosen for analysis

- SQL
- Python

3

Database Design & Data Transformation

4

Cleaned & Filtered Script

5

Insights & Results

# NYC Dataset selected - Data Acquisition, Transformation and Examination:

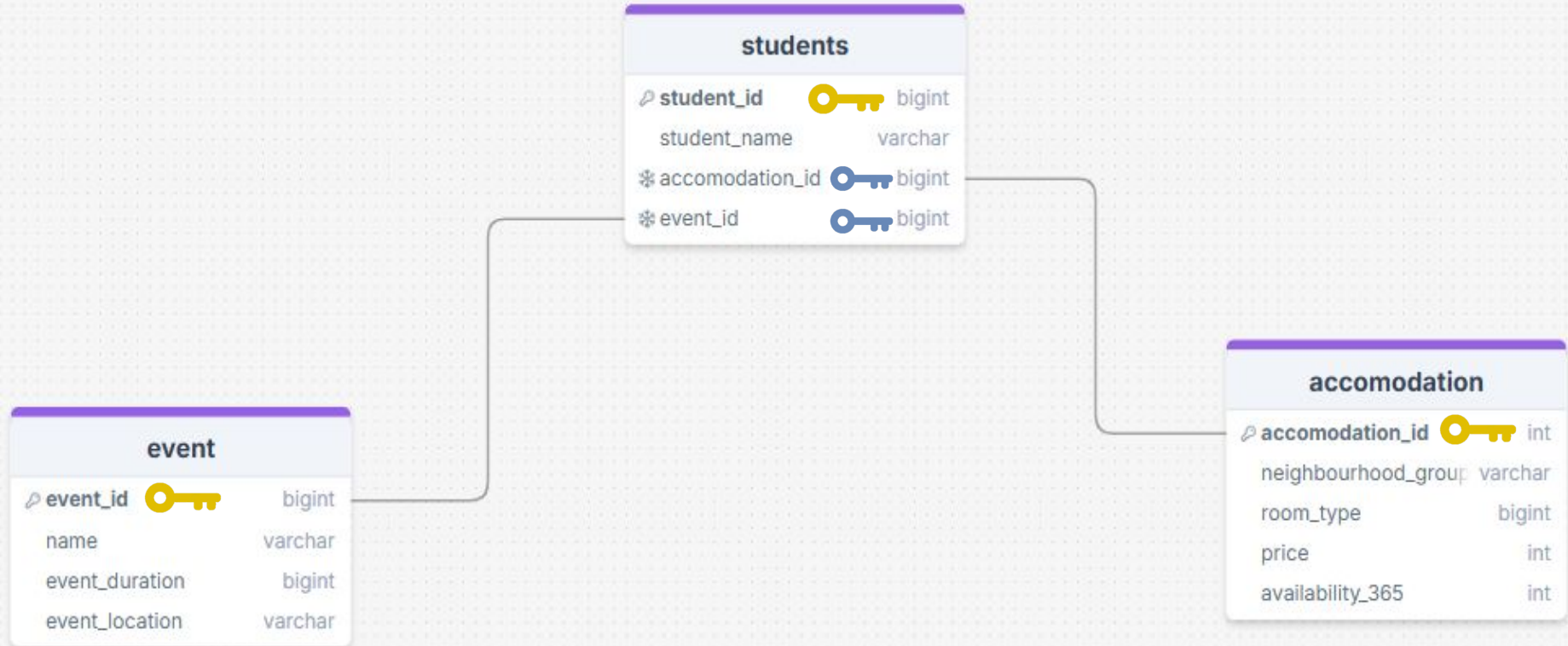
**A university office is tasked with trying to find the closest and most cost effective AirBnB in NYC for a group of students attending events in the city.**

Topics covered:

- ★ Highlight the business problem and the hypotheses that guided our approach.
- ★ Highlight the primary sources of data and any complementary datasets.
- ★ Data acquisition, transformation and examination process
- ★ Discuss challenges faced during the data sourcing and integration.
- ★ Briefly describe how the supplemental data aligns with our primary dataset.

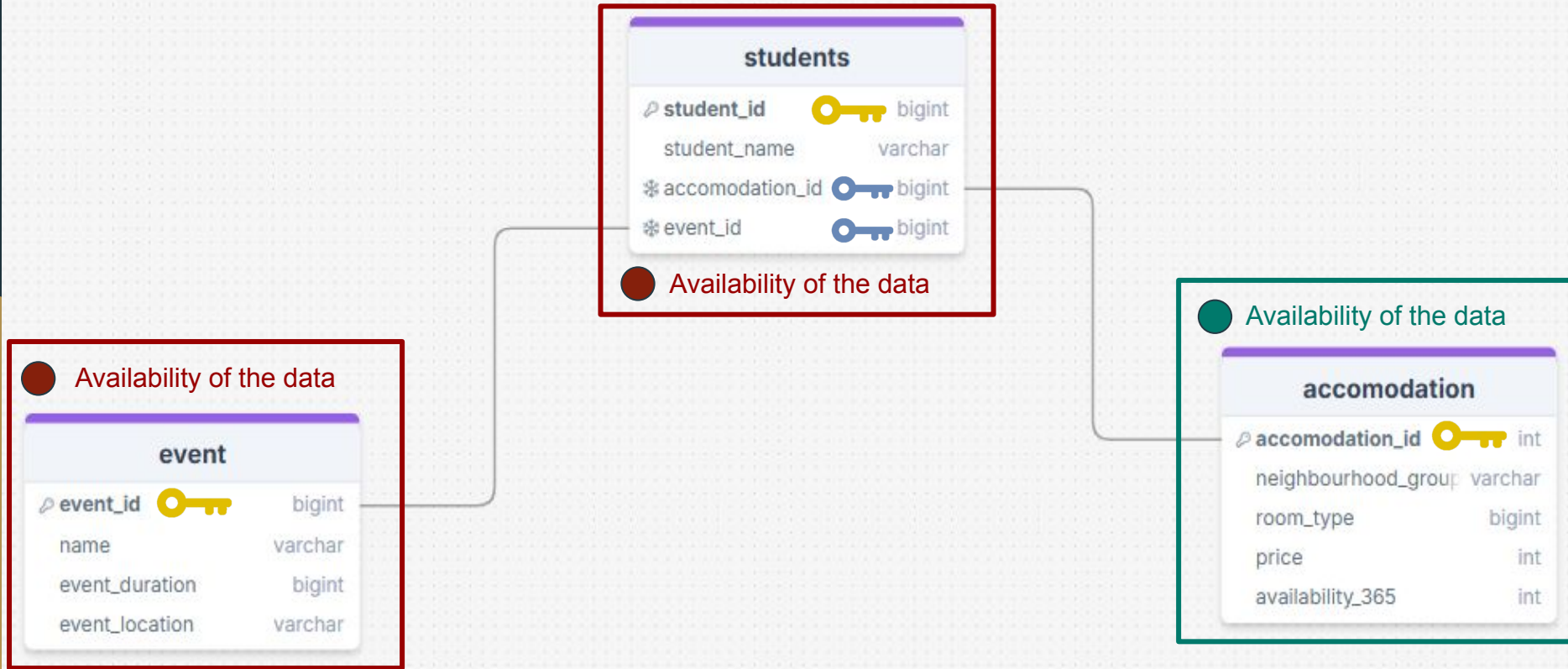
# Database Design & Data Transformation:

## StudentHub: NYC Rentals : Entity–Relationship Diagram (ERD) - project



# Database Design & Data Transformation:

## StudentHub: NYC Rentals : Entity–Relationship Diagram (ERD) - project



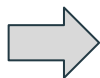
# Database Design & Data Transformation:

## StudentHub: NYC Rentals : Entity–Relationship Diagram (ERD) - realisation

students				
student_id	student_name	accommodation_id	event_id	
1	Mary Smith	33506	1	
2	John Holland	2861	2	
3	Dillon Robbins	33226	3	
4	Michael Beck	33226	4	
5	Rachel Adams	34447	5	

event			
event_id	name	event_duration	event_location
1	Conference 1	14	Brooklyn, New York
2	Conference 2	74	Manhattan, New York
3	Conference 3	36	Queens, New York
4	Conference 4	45	Staten Island, New York
5	Conference 5	68	Bronx, New York

accommodation						
accommodation_id	neighbourhood_group	room_type	price	availability_365		
1	Brooklyn	Private room	149	365		
2	Manhattan	Entire home/apt	225	355		
3	Manhattan	Private room	150	365		
4	Brooklyn	Entire home/apt	89	194		
5	Manhattan	Entire home/apt	80	0		
6	Manhattan	Entire home/apt	300	139		



ny_df															
id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
0	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365
1	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355
2	THE VILLAGE OF HARLEM, NEW YORK	4632	Elisabeth	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN	NaN	1	365
3	Cozy Entire Floor of Brownstone	4569	Lisa/Joanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194
4	Entire Apt: Spacious Studio w/ central park	7192	Laura	Manhattan	East Harlem	40.79651	-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
48890	Charming renovated room in Brooklyn	822,024	Sabrina	Brooklyn	Shuysant	40.67501	-73.97501	Private room	40	70	0	NaN	NaN	2	9
48891	Affordable rooms in Bushwick/East Williamsburg	65706		Brooklyn				Private room	70	70	0	NaN	NaN	2	36
48892	Sunny Studio at historical Neighborhood	23492952	Igor & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN	NaN	1	27
48893	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	NaN	NaN	6	2
48894	Trendy duplex in the very heart of Hell's Kitchen	68118614	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98953	Private room	90	7	0	NaN	NaN	1	23

48895 rows = 16 columns

events\_df.csv

students\_df.csv

```
[38]: # Create DataFrame
events_df = pd.DataFrame(events)

[39]: events_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 4 columns):
# Column      Non-Null Count  Dtype
---  ---  ---
0 event_id      5 non-null      int64
1 event_name    5 non-null      object
2 event_duration 5 non-null      int64
3 event_location 5 non-null      object
dtypes: int64(2), object(2)
memory usage: 292.0+ bytes

[41]: # Show table
print(events_df)

event_id  event_name  event_duration  event_location
0         1  Conference  14              Brooklyn, New York
1         2  Conference  74              Manhattan, New York
2         3  Conference  36              Queens, New York
3         4  Conference  45              Staten Island, New York
4         5  Conference  68              Bronx, New York
```

```
[26]: ### students.csv file ###

[27]: # Students data
students = [{"student_id": 1,
             'student_name': 'Mary Smith'},
            {'student_id': 2,
             'student_name': 'John Holland'},
            {'student_id': 3,
             'student_name': 'Dillon Robbins'},
            {'student_id': 4,
             'student_name': 'Michael Beck'},
            {'student_id': 5,
             'student_name': 'Rachel Adams'}]

[28]: students

[28]: [{"student_id": 1, 'student_name': 'Mary Smith'},
      {'student_id': 2, 'student_name': 'John Holland'},
      {'student_id': 3, 'student_name': 'Dillon Robbins'},
      {'student_id': 4, 'student_name': 'Michael Beck'},
      {'student_id': 5, 'student_name': 'Rachel Adams'}]

[29]: students_df = pd.DataFrame(students)
```

ny\_df\_cleaned\_final.csv

```
[6]: ny_df_cleaned_final

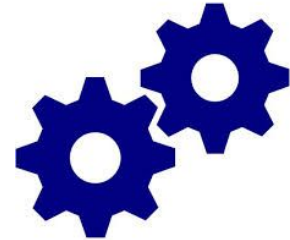
[6]: neighbourhood_group  room_type  price  availability_365
0 Brooklyn Private room 149 365
1 Manhattan Entire home/apt 225 355
2 Manhattan Private room 150 365
3 Brooklyn Entire home/apt 89 194
4 Manhattan Entire home/apt 80 0
... ..
48890 Brooklyn Private room 40 9
48891 Brooklyn Private room 70 36
48892 Manhattan Entire home/apt 115 27
48893 Manhattan Shared room 55 2
48894 Manhattan Private room 90 23

48895 rows x 4 columns
```

# Database Design & Data Transformation:

StudentHub: NYC Rentals : Entity–Relationship Diagram (ERD) - realisation

MySQL



students				
	student_id	student_name	accommodation_id	event_id
▶	1	Mary Smith	33506	1
	2	John Holland	2861	2
	3	Dillon Robbins	33226	3
	4	Michael Beck	NULL	4
	5	Rachel Adams	34447	5
*	NULL	NULL	NULL	NULL

event				
	event_id	name	event_duration	event_location
▶	1	Conference 1	14	Brooklyn, New York
	2	Conference 2	74	Manhattan, New York
	3	Conference 3	36	Queens, New York
	4	Conference 4	45	Staten Island, New York
	5	Conference 5	68	Bronx, New York
*	NULL	NULL	NULL	NULL

# accommodation

	accommodation_id	neighbourhood_group	room_type	price	availability_365
▶	1	Brooklyn	Private room	149	365
	2	Manhattan	Entire home_apt	225	355
	3	Manhattan	Private room	150	365
	4	Brooklyn	Entire home_apt	89	194
	5	Manhattan	Entire home_apt	80	0
	6	Manhattan	Entire home_apt	200	129
	-	-	-	-	-

# Python cleaning and SQL database creation:

## Jupyter Notebook

Cleaning of the dataset we chose (Airbnb NY) using python and pandas

Generation of three CSV files

IMPORTED

## SQL Database

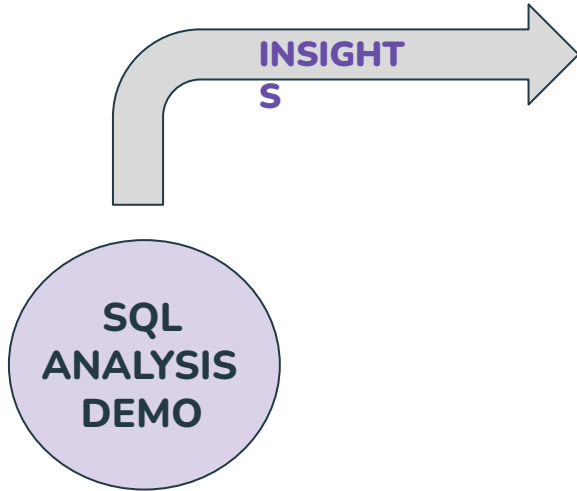
Created database

Created the tables

Imported CSV files to populate the tables



# SQL Insights & Advanced Analysis:



- No accommodation in the dataset that had enough availability period located in Staten Island
- We could determine the cheapest option in the neighbourhood that had enough and the higher availability for each student, considering the event to attend
- We decided that Michael Beck will stay at the Airbnb in Manhattan with John Holland, as it is an entire home apartment and John's event will take longer than Michael's event.

# Conclusions & Business Implications:

## Best Accomodation:

1 Each student is linked to an event and assigned an accommodation based on location, availability, budget and room type

## Condition

Price is more than 0

2 If multiple options exist, one with the highest availability is chosen.

## Criteria

3 Accomodation is in the same neighbourhood, as the event.

Available at least for the whole event duration

The students are allotted accommodations such as Entire home apartments or Private rooms, ensuring the most cost-effective and suitable options for their respective events. However, since accommodation was not available for Student 4, they will need to adjust by sharing or staying near other students in nearby locations such as Manhattan or Brooklyn.

DATA-FT-Sept-2025

MiniProject2 - SQL: From Data to Insight

NYC Airbnb for STUDENTS

StudentHub: NYC Rentals

**Thank You !**

Kinga, Julia, Joma, Priyanka