# Mini Project – SQL: From Data to Insight

### Airbnb NYC Case Study

## 👥 Team Members

- Ricardo Castanheira

- Irma Fernandez

- Janna Julian

- Luis Pablo

---

## 🎯 Project Overview

This project explores the \*\*business opportunity\*\* of opening a new Airbnb listing in New York City.

We use the \*\*NYC Airbnb dataset (Kaggle)\*\* to analyze pricing, availability, and room types in different boroughs.

---

## 📅 Project Plan (Daily Tasks)

### Day 1 – Project Initiation & Data Selection

- Selected dataset: New York City Airbnb Open Data (Kaggle)

- Defined business opportunity: \*Evaluate potential for opening an Airbnb listing in NYC\*

- Hypotheses:

1. Manhattan is the most profitable area.

2. Entire homes/apartments are more in demand.

3. 365-day availability correlates with revenue potential.

- Created project board (Trello/Jira) for task distribution.

### Day 2 – Data Examination & Schema Design

- Explored dataset and cleaned data (removed duplicates, capped outliers).

- Designed \*\*Entity Relationship Diagram (ERD)\*\* with 4 tables:

- `neighbourhood\_group`

- `neighbourhoods`

- `rooms\_type`

- `listings`

- Created MySQL database schema.

- Loaded cleaned data into normalized tables.

### Day 3 – SQL Queries & Analysis

- Wrote SQL queries to answer business questions:

- Average price by borough.

- Most common room type per borough.

- Minimum nights average.

- Availability buckets vs. price.

- Top 5 most expensive neighborhoods.

- Validated results with joins, group by, and subqueries.

### Day 4 – Analysis & Visualization

- Used Python (Pandas, Matplotlib/Seaborn) to visualize insights:

- Bar chart – Average price by borough.

- Pie chart – Room type distribution.

- Scatterplot – Availability vs. Price.

- Compiled findings into final report (Jupyter Notebook).

- Designed project presentation (Google Slides/Canva).

---

## 📊 Findings

- Manhattan has the highest average prices, but strong competition.

- Brooklyn offers more affordability with consistent demand.

- Entire apartments dominate the market, while private rooms remain niche.

---

## 💡 Business Recommendation

- Opening a listing is feasible in \*\*Brooklyn\*\*, offering balance between affordability and demand.

- Recommended setup: \*\*Entire Apartment\*\* at a price slightly below borough average.

- Availability should be flexible, not necessarily 365 days, to optimize operations.

## 📂 Repository Structure

sql-database/

├─ README.md

├─ data/

│ └─ nyc\_airbnb.csv

├─ sql/

│ ├─ 01\_create\_staging.sql

│ ├─ 02\_schema\_erd\_4tables.sql

│ ├─ 03\_load\_dimensions.sql

│ ├─ 04\_load\_listings.sql

│ └─ 05\_queries\_insights.sql

├─ erd/

│ └─ airbnb\_erd.png

├─ notebooks/

│ └─ 01\_report.ipynb (visualizaciones finales)

└─ src/

└─ etl.py (funciones limpias de carga/limpieza opcional)