Data Engineer Case

In this data engineer case, you are expected to investigate and explore different datasets. The case is quite open so you are free interpret the task as you wish. We do not expect you to use more than 4-6 hours, the purpose of the case is to use it as a basis for a technical interview.

Datasets:

1. Airbnb\_data\_New\_York.csv
2. Property\_sales\_data\_New\_York.csv
3. Weather data.xslx

The **Property\_sales\_data\_New\_York** dataset contains a column “BOROUGH”. This column includes a digit code for the borough the property is located. These are: Manhattan (1), Bronx (2), Brooklyn (3), Queens (4), and Staten Island (5).

Tasks:

1. Present some of the key findings of the datasets above.

Some suggestions for things you can try to find out:

* + How many properties are sold with zero price?
  + What is the average price for each neighborhood to rent an Airbnb apartment?
  + What are the number of sales of properties in each neighborhood in NY?
  + Are sales of properties in NY impacted by the weather temperatures?

1. If you were supposed to buy an apartment in New York, what would be your approach to find the best investment if you plan to use it for Airbnb?