LOW LEVEL DESIGN DOCUMENT

(Airbnb Data Analysis)

Akshay Salve

VERSION: 1.0

DATED: 26/06/2022

Document Version Control:

Airbnb Data Analysis – Business Intelligence Project

Version	Date	Author	Change
1.0	27/06/2022	Akshay R. Salve	First version of LLD

Abstract

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.

This dataset provides huge data on based on the survey conducted in July 2017.Based on this information the ultimate goal is to find the insights highlighting key indicators and metrics that influence the travel market.

Contents:

1. Introduction

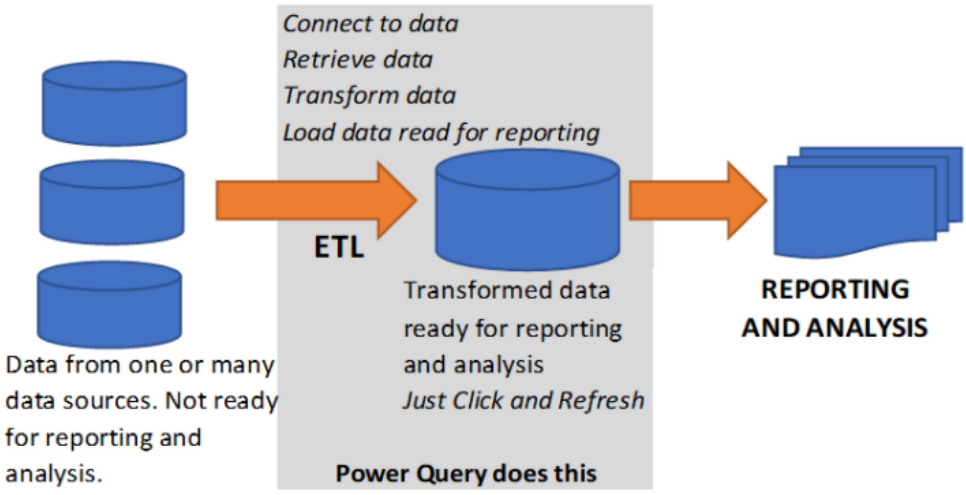
1.1. Why this Low-Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Bank Marketing Campaign Analysis. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture:



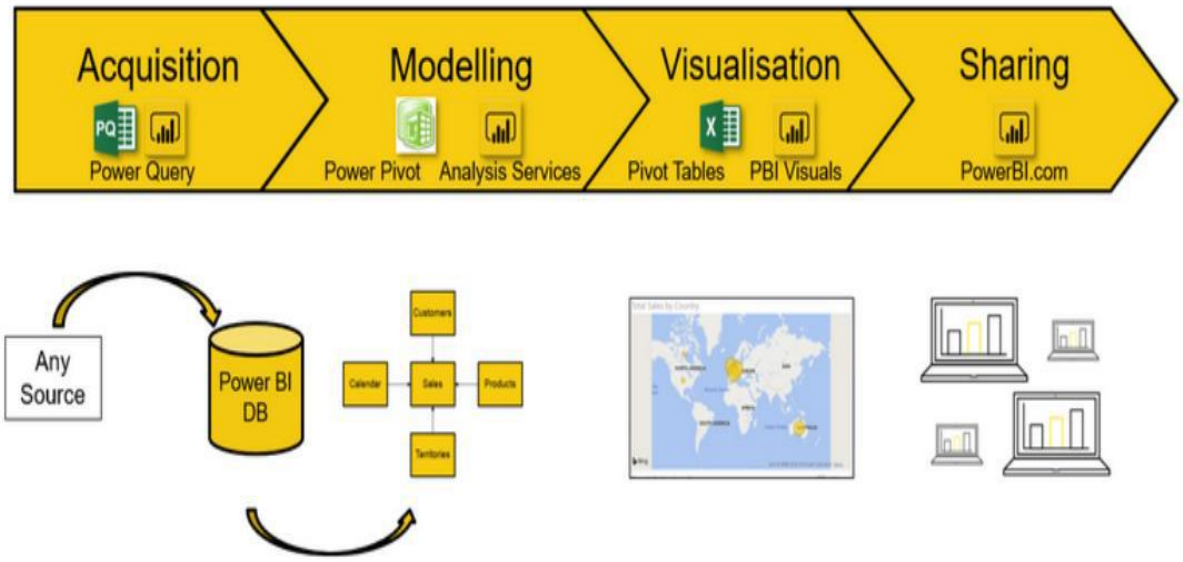
ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

Based on the results of ETL, companies also make business decisions, which can have repercussions later.

- If ETL is not done properly then it can damage the business a lot in many ways such as loss of client which we are working for, the decision making will go completely wrong and many more issues.
- If done well, it may improve the efficacy of everything we do next.

Below are following steps to follow for ETL:

1. Data Sourcing
2. Data Cleaning
3. Data Modelling
4. Data Visualization



3. Architecture Description:

3.1 Data Sourcing:

The dataset is in csv (comma separated values) format. MS Excel is used to load the data.

Citation Request:

This dataset is provided by the client.

1.Title - Airbnb_prices.csv.

3.2. Data Overview –

- The data includes single csv file with all examples, ordered by date (22 July 2017 to 23 July 2017.
- The Number of Instances – 18724 for Airbnb_price.csv
- Number of Attributes – 19 Attributes

3.3. Data Description –

- room_id – Ids for rooms [10176931, 8935871, 14011697, ..., 19859427, 17132164, 7605782]
- survey_id – Id od survey conducted [1476]
- host_id – Ids for hosts [49180562, 46718394, 10346595, ..., 139135665, 1501422, 29724632]
- room_type – Type of rooms ['Shared room', 'Entire home/apt', 'Private room']
- country – empty
- city – Amsterdam
- borough – empty
- neighbourhood – Names of neighbourhoods offering services ['De Pijp / Rivierenbuurt', 'Centrum West', 'Watergraafsmeer', 'De Baarsjes / Oud West', 'Oostelijk Havengebied / Indische Buurt', 'Westerpark', 'Oud Oost', 'Centrum Oost', 'Geuzenveld / Slotermeer', 'Buitenveldert / Zuidas', 'Oud Noord', 'Bos en Lommer', 'Slotervaart', 'Bijlmer Centrum', 'Ijburg / Eiland Zeeburg', 'Noord-West / Noord-Midden', 'De Aker / Nieuw Sloten', 'Noord West', 'Bijlmer Oost', 'Osdorp', 'Noord Oost', 'Gaasperdam / Driemond', 'Westpoort']
- reviews – numeric
- overall_satisfaction – numeric
- accommodates – numeric
- bedrooms – numeric
- bathrooms – empty
- price – numeric
- mainstay – empty
- name – names of stays
- last_modified – Date of survey
- latitude – North-South coordinates of the room location

- ❖ longitude – East-West coordinates of the room location
- ❖ location – Exact location of the room

3.4 Data loading in Power BI Query Editor

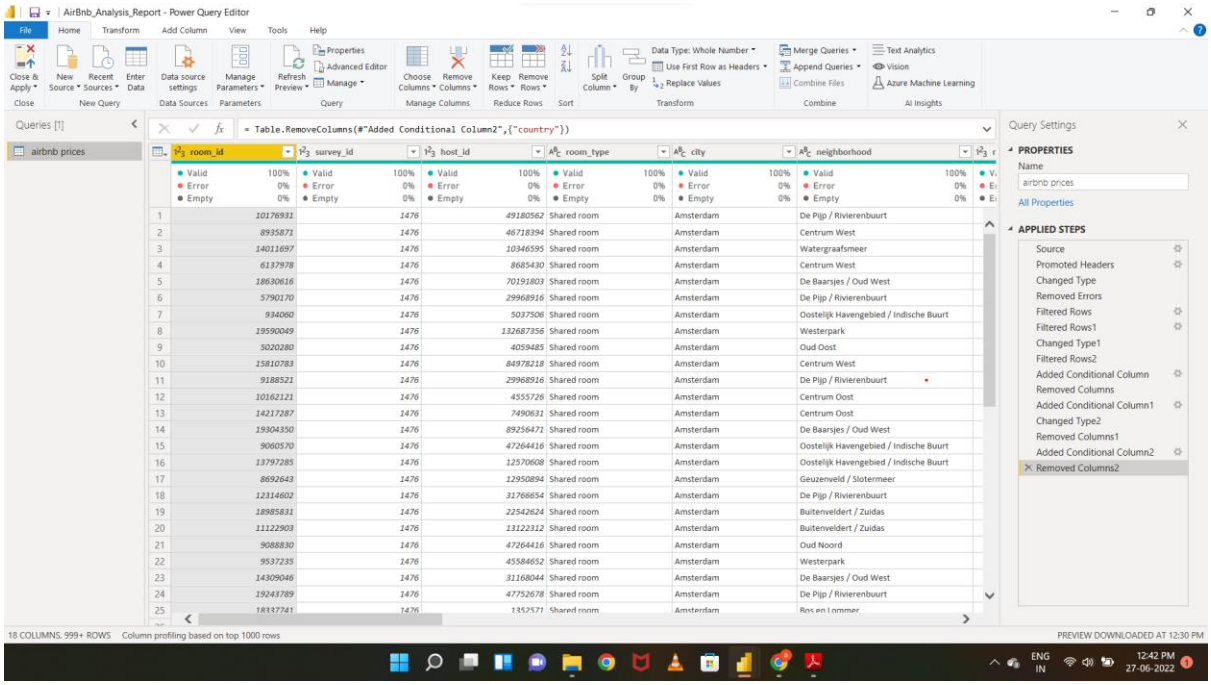
Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data verse, and more with the following characteristics:

❖ There can be multiple rows and columns in the data.

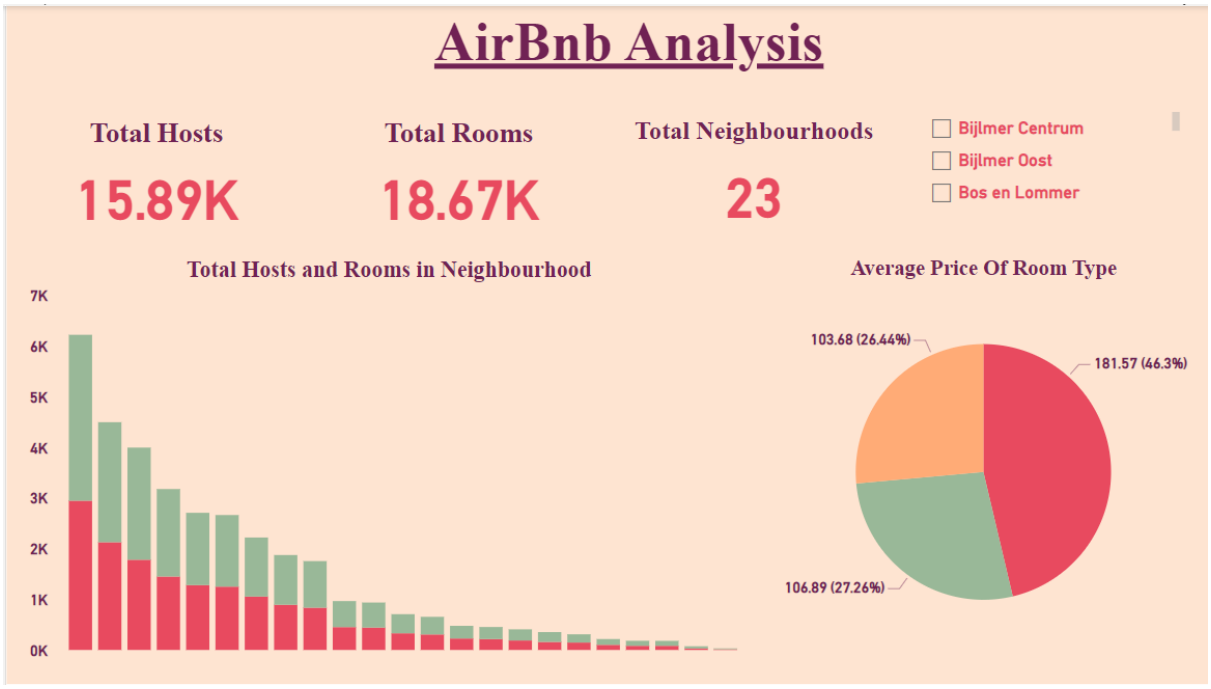
❖ Each row represents a sample of data,

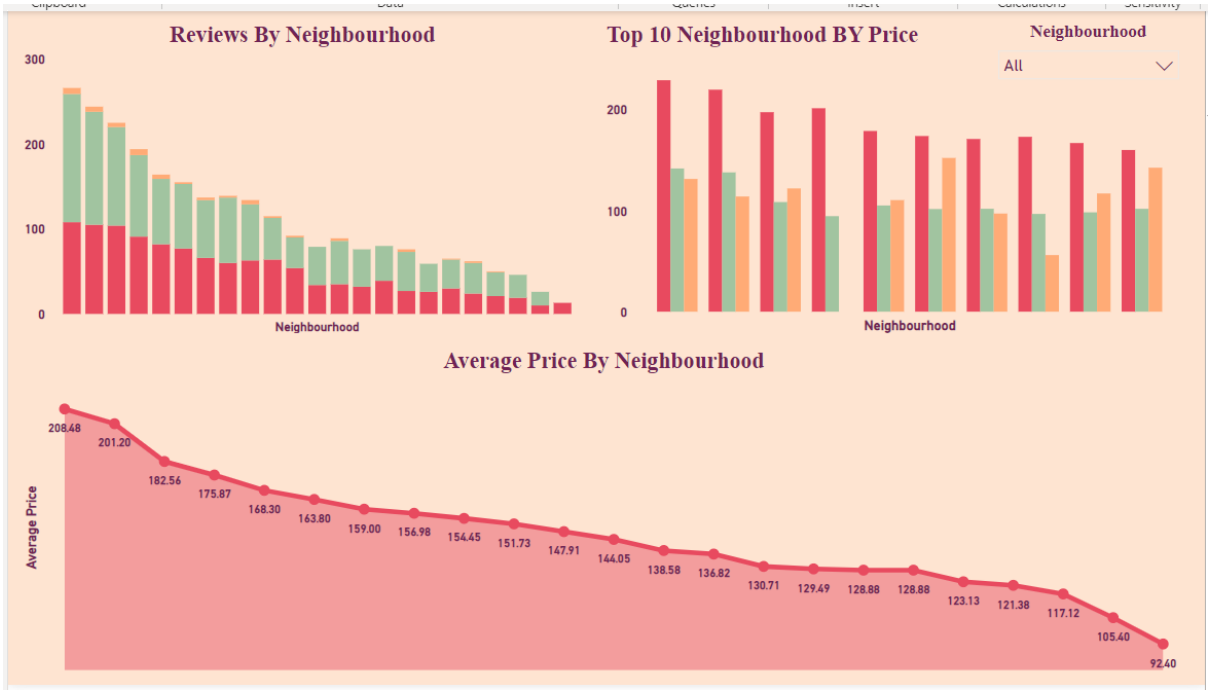
❖ Each column contains a different variable that describes the samples (rows).

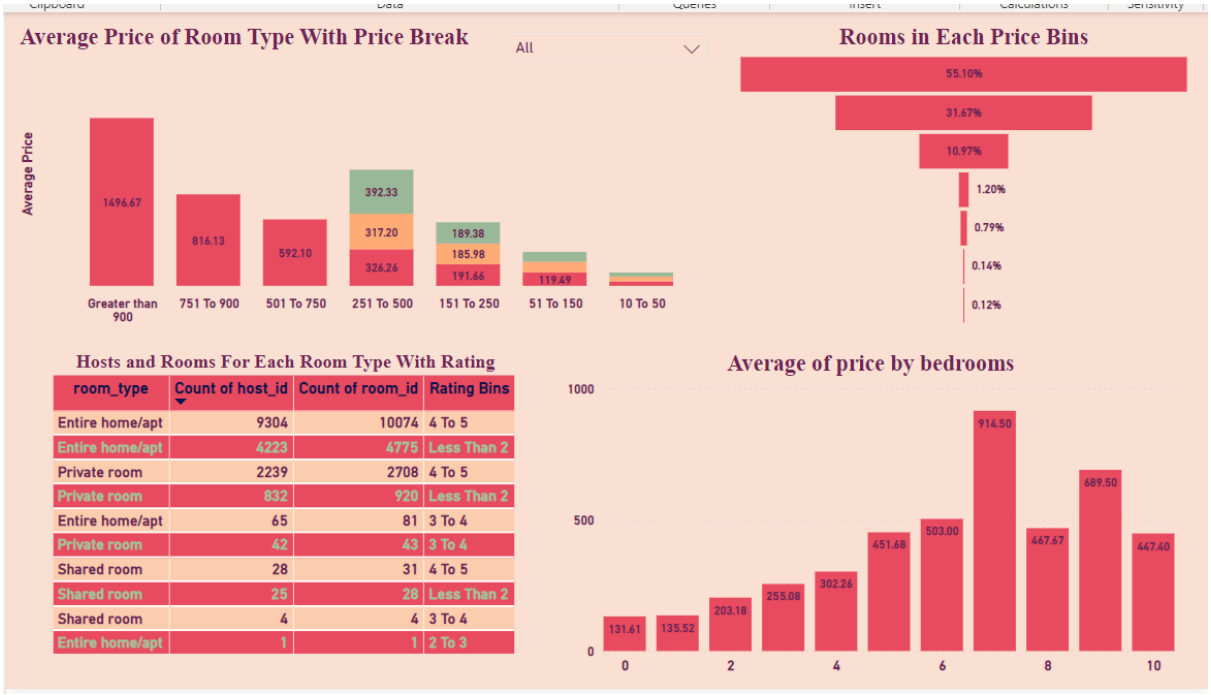
❖ The data in every column can be a different type of data – e.g. numbers, strings, dates, Boolean etc.



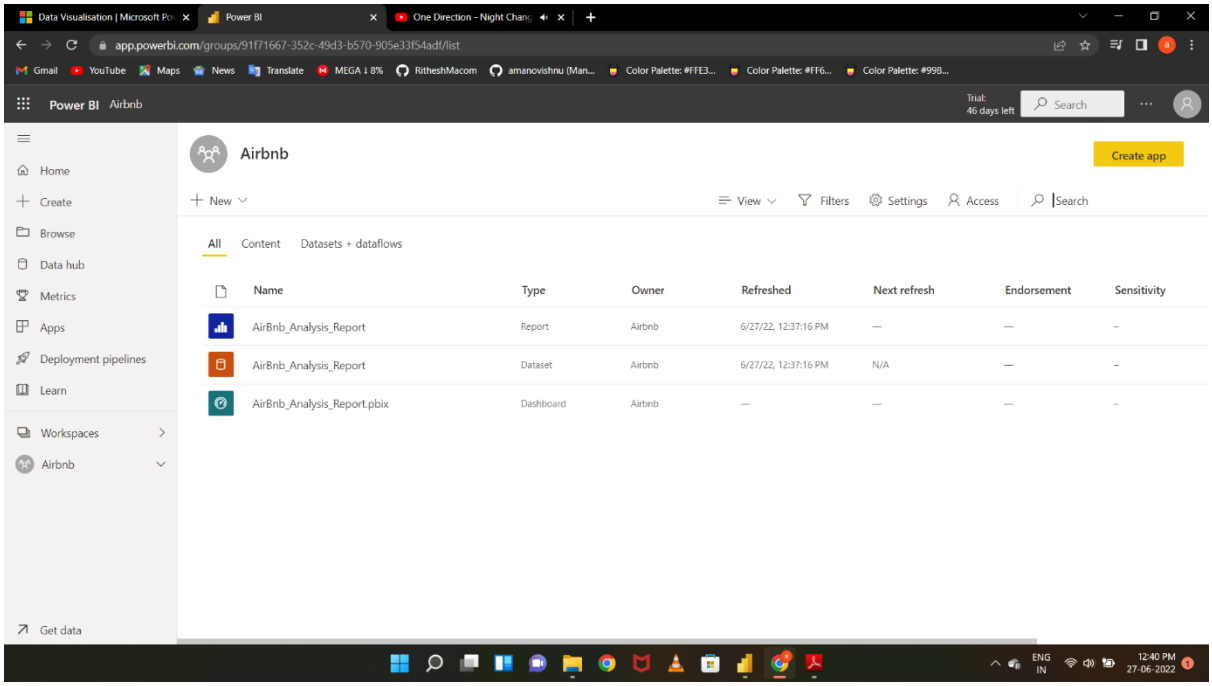
3.5 Data to Insights through Visualizations and Excel Data Analysis







3. Deployment to Power BI Service



Low Level Design (LLD)