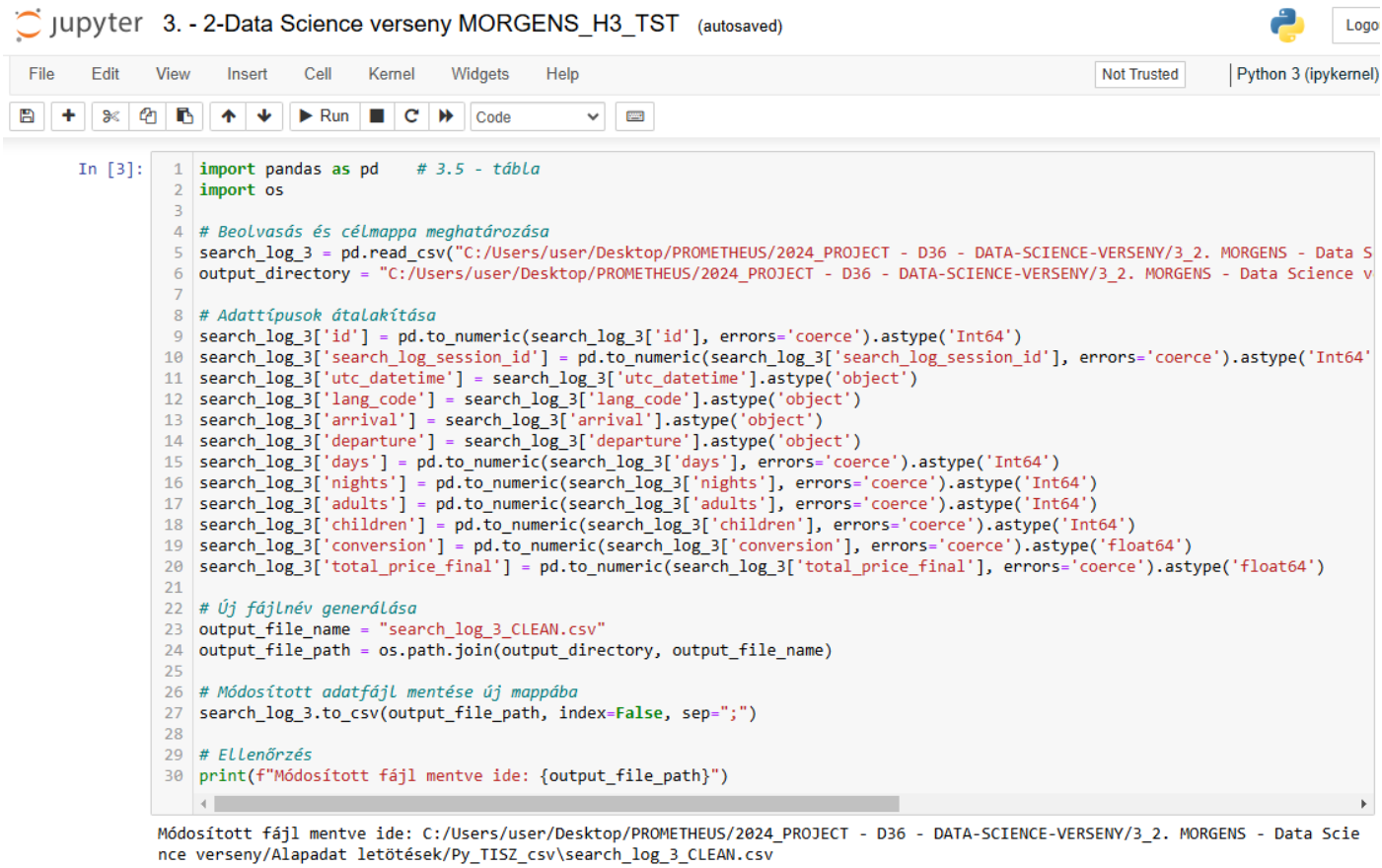
MORGENS Hotel Management Business Efficiency Improvement

1. Introduction:

The objective of this project is to uncover hidden insights in business-related questions by analyzing the correlation between the starting point and the target outcome. The starting point is the cleaned and prepared source data, while the target outcome is the maximized monthly revenue. Beyond answering business questions, this approach also reveals potential opportunities.

2. Data Sources:

- **11 CSV** files containing data on hotel user activities.
- **3 IPYNB** files detailing the data preparation methodology for each hotel.
- **Data transformation:**
    - Began in Python (Jupyter Notebook environment).:

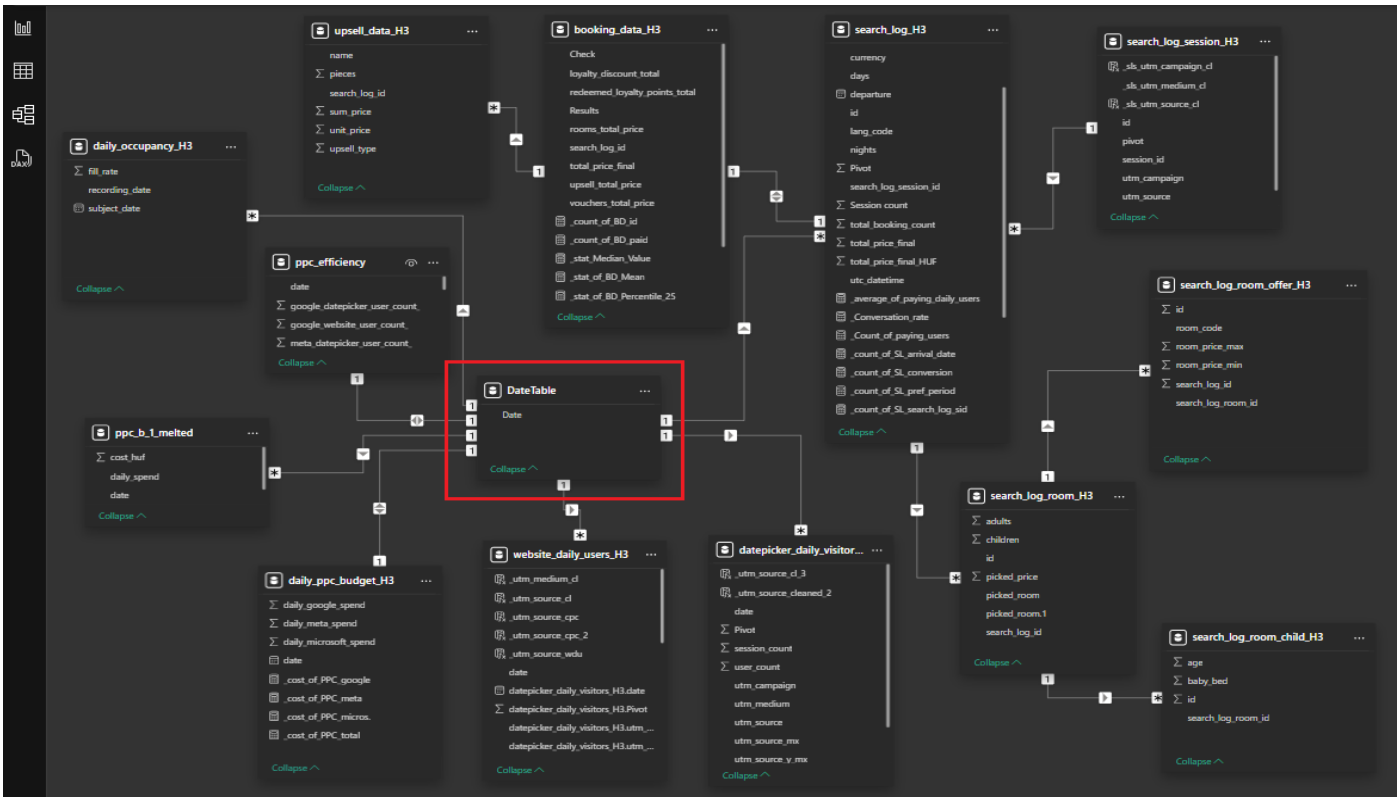


    - The process was finalized in Power Query, which will be further detailed later.

3. Data Model:

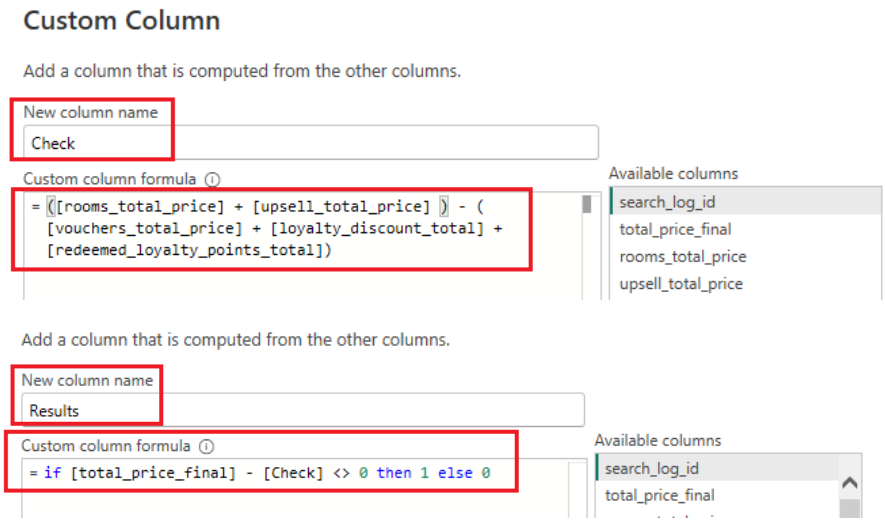
The data model was designed to integrate the prepared source tables into a single data warehouse, where bidirectional and unidirectional relationships facilitate the discovery of insights and actionable recommendations.

The complete data model structure is illustrated in the image below:



- **Table Relationships:** The first step was to create a bridge table (highlighted in red above) due to duplicate date values in the connected data tables. Therefore, a separate date table with unique values was introduced to ensure accurate aggregation and time-based analysis (Star Schema).
- Key Tables: Two primary table relationship methods were used in this model:
    - **One-to-Many** (1:*) – When only one table contains unique values (e.g., image below 2, 3, 4).
    - **Many-to-One** (*:1) – The reverse of the One-to-Many relationship (e.g., image below 2, 3, 4).
    - **One-to-One** (1:1) – When all tables contain only unique values (e.g., image below 5, 6, 7).

- **Data Cleaning & Normalization:** The primary data cleaning and preparation were conducted in Python (Jupyter Notebook), but additional transformations were needed during the process. These were handled in Power Query within Power BI. The most common operations included:
    - Handling NaN, empty, zero values, and character errors
    - Trimming white spaces
    - Converting text to numeric values using the Replace Value function.
- **Feature Engineering:**
    - *Validation Purposes:*
        - During the EDA, inconsistencies were found in the booking_data table.
        - A Custom Column M-function was created in Power Query to check and validate values:



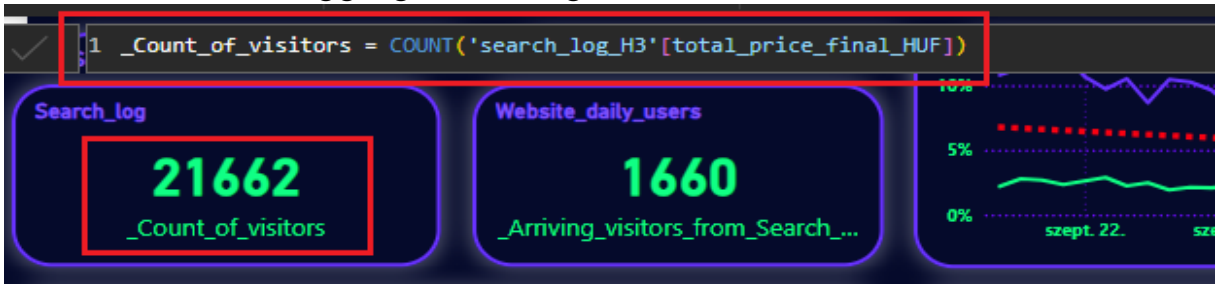
        - An additional column was added to verify whether Check and total_final_price column match. If not, it outputs 1.

4. DAX Formulas & Key Metrics:

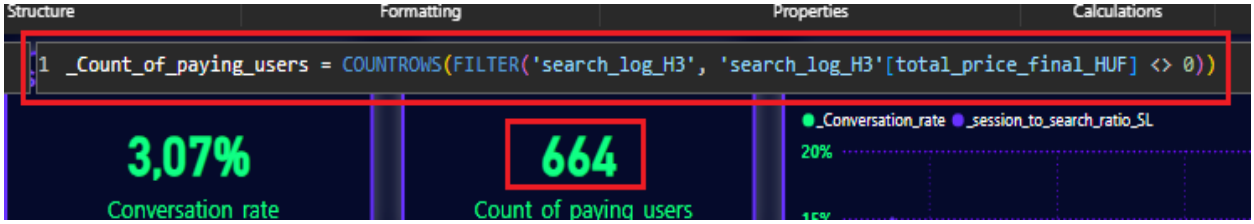
The most important measures included:

- **Conversion rates:**

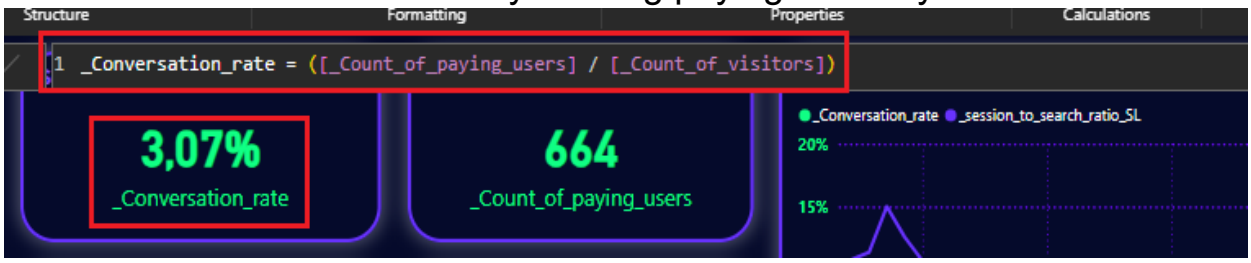
  Visitor count was aggregated using the COUNT DAX function:



Paying customers were counted using the COUNTROWS FILTER DAX function:



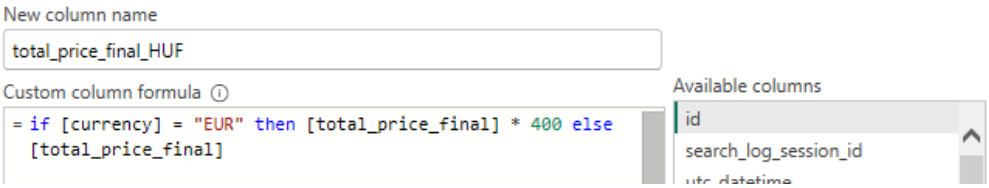
Conversion rate was derived by dividing paying users by total visitors:



- Revenue
- Advertising efficiency

Below are key calculations:

- Currency Conversion with M-function
  - A currency conversion M function was implemented to standardize values across different currencies.



- Descriptive Statistics in Power BI:
  - Unlike Python, where statistics can be generated using **booking_1.describe(),** Power BI requires individual calculations.

```
1 _stat_of_BD_Mean = AVERAGE('booking_data_H3'[total_price_final])
1 _stat_of_BD_Median = MEDIAN('booking_data_H3'[total_price_final])
1 _stat_of_BD_Percentile_25 = PERCENTILEX.INC(ALL('booking_data_H3'), 'booking_data_H3'[total_price_final], 0.25)
1 _stat_of_BD_Percentile_75 = PERCENTILEX.INC(ALL('booking_data_H3'), 'booking_data_H3'[total_price_final], 0.75)
1 _stat_of_BD_Std_Dev = STDEV.P('booking_data_H3'[total_price_final])
```

- Text Categorization & Data Manipulation in DAX:

```
_utm_source_cl_3 =
SWITCH(
    TRUE(),
    'datepicker_daily_visitors_H3'[_utm_source_cleaned_2] IN {"facebook", "instagram"}, "meta",
    'datepicker_daily_visitors_H3'[_utm_source_cleaned_2] IN {"bing", "edge"}, "microsoft",
    'datepicker_daily_visitors_H3'[_utm_source_cleaned_2] = "google", "google",
    'datepicker_daily_visitors_H3'[_utm_source_cleaned_2] = "direct", "direct",
    'datepicker_daily_visitors_H3'[_utm_source_cleaned_2] = "not set", "na",
    "other"
)
```

- Multi-Condition Categorization in DAX:

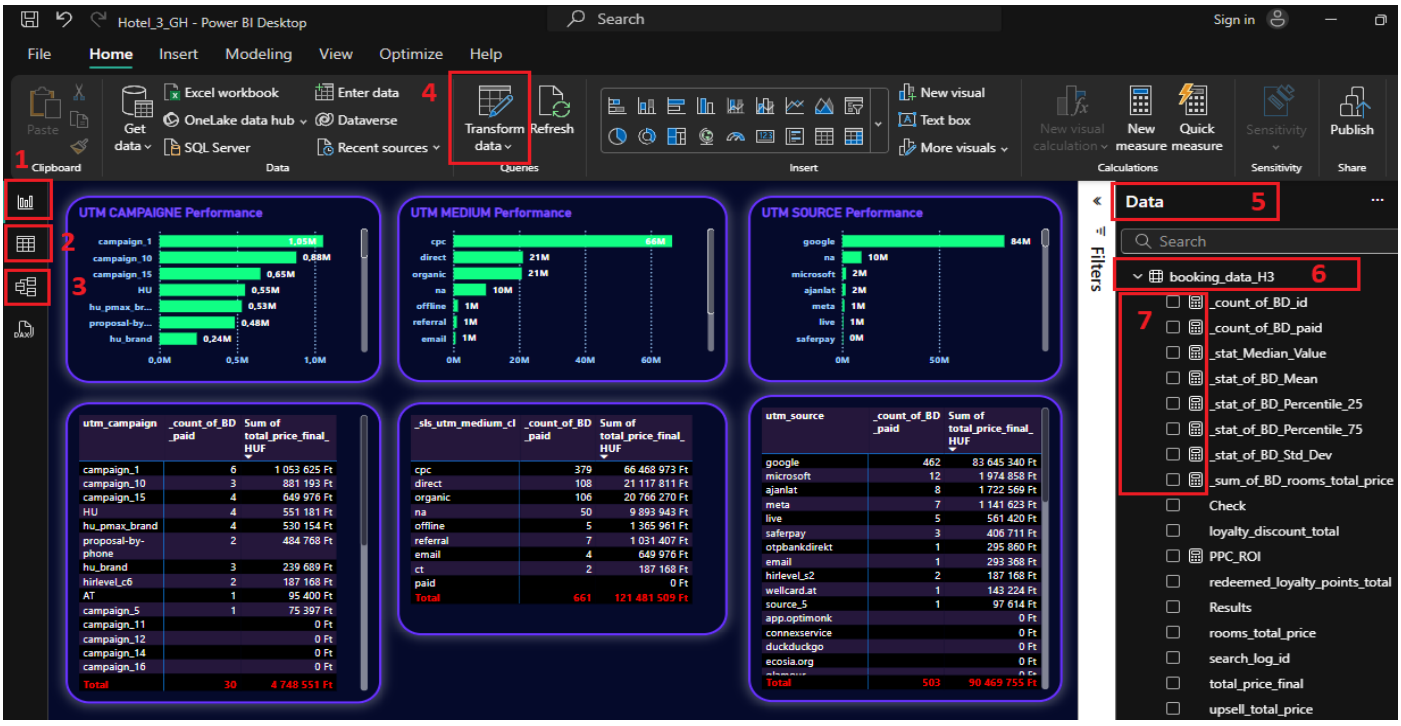
```
_utm_source_cl =
SWITCH(
    TRUE(),
    SEARCH("facebook", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "facebook",
    SEARCH("instagram", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "instagram",
    SEARCH("bing", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "bing",
    SEARCH("google", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "google",
    SEARCH("direct", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "direct",
    SEARCH("offer", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "offer",
    SEARCH("not set", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "not set",
    SEARCH("edge", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "edge",
    SEARCH("phone", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "phone",
    SEARCH("otpbankdirekt.hu", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "otpbankdirekt.hu",
    SEARCH("kereso.startla.hu", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "kereso.startla.hu",
    SEARCH("loyalty", 'website_daily_users_H3'[utm_source], 1, 0) > 0, "loyalty",
    "other" -- ha egyik feltétel sem teljesül
)
```

5. Visualization & Report Structure:

Power BI consists of two main components:

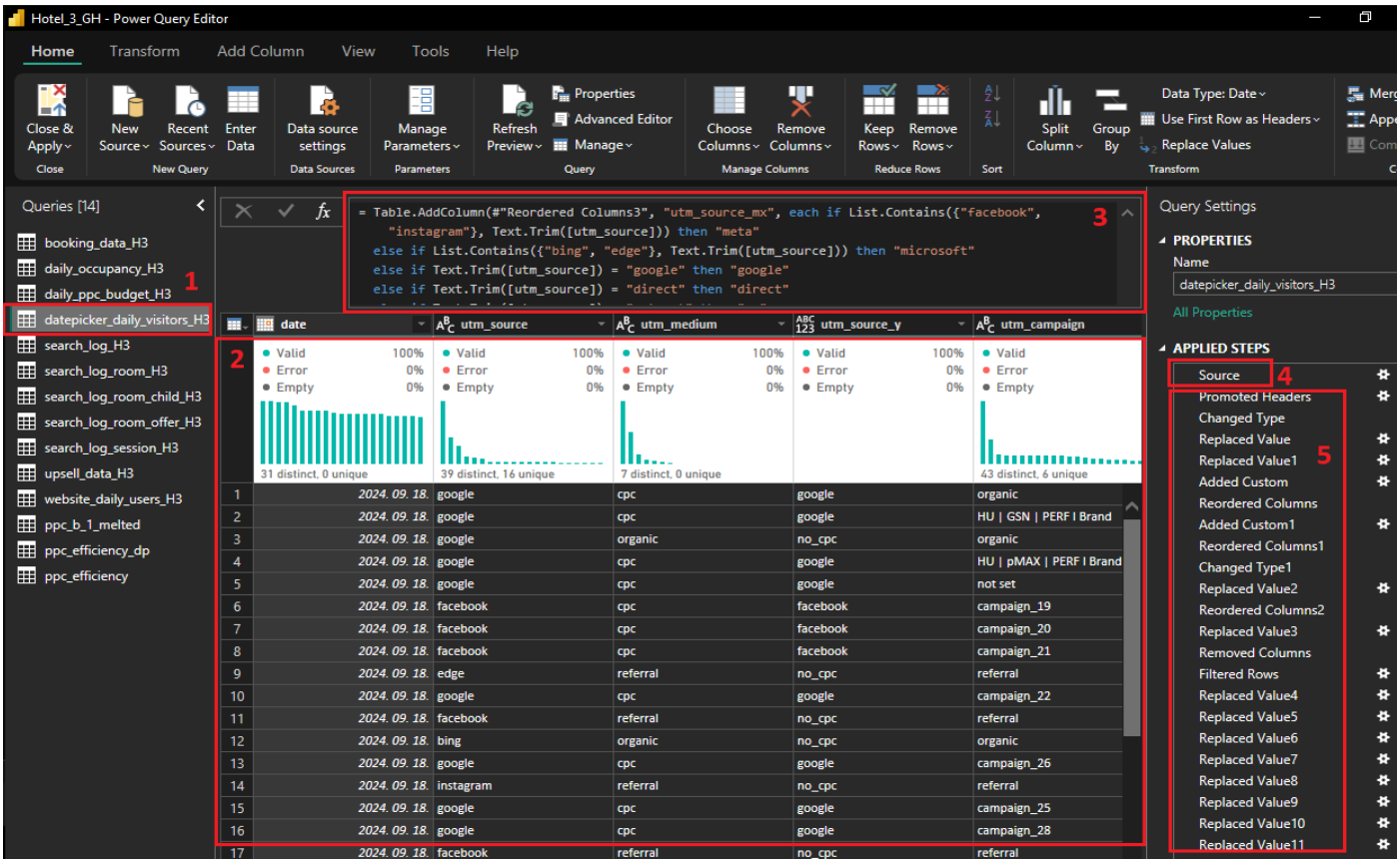
1. **Dashboard** – Used for data visualization.
2. **Power Query Editor** – Used for data preparation and cleaning.

Dashboard Structure: The project's Dashboard consists of three key sections (as seen on the left in the image below):



- **Report View (1):** Used for visualizing charts, pivot tables, matrices, and writing DAX formulas.
- **Table View (2):** Displays table contents without entering the editor.
- **Model View (3):** Shows the data model relationships, the core structure of the system.
- **Power Query (4):** Used for advanced data transformations.
- **Data Tables & Columns (5 & 6):** Displays available datasets and their column names.
- **DAX Generated Measures (7):** Represented by a small calculator icon, best stored at the top of the list for accessibility.

Power Query Structure: Power Query is responsible for data preparation and consists of the following sections:



- **(1) Loaded Tables:** Displays imported datasets.
- **(2) Data Table:** Allows for additional data cleaning.
- **(3) Formula Bar:** Uses the M language for transformations.
- **(4) Transformation Steps:** Each step is logged and sequentially processed.
- **(5) Final Filtered Data:** Ensures only relevant values are included.

6. Automated Model Refresh:

The final model enables seamless integration with new monthly data. By simply pressing the refresh button, the updated report is generated in less than 30 seconds.

Additionally, if both the source data and Power BI model were hosted on a server, an automatic update could be scheduled (e.g., every day at 12:05 AM) without manual intervention.