Name :- Shruti More

Module 1: Data Preparation & Guest Stay Analysis

Module Overview

Module 1 focuses on building a strong data foundation for the hotel analytics project. In this module, raw booking data is cleaned, transformed, and enriched to analyze guest stay behavior, duration patterns, and branch-wise stay distribution.

This module ensures that all subsequent analyses are based on accurate, well-structured, and meaningful data.

Data Tables Used

1. HotelBookings (Fact Table)

Contains detailed booking-level information:

- BookingID
- CustomerID
- RoomID
- BranchID
- CheckInDate
- CheckOutDate
- Duration of Stay
- RoomRate
- TotalAmount
- Purpose of Stay
- Stay_Type

2. Customer (Dimension Table)

- CustomerID
- Age
- Gender
- Country
- LoyaltyTier

3. Room (Dimension Table)

- RoomID
- RoomType
- BedType
- Capacity
- PricePerNight
- Room_Category

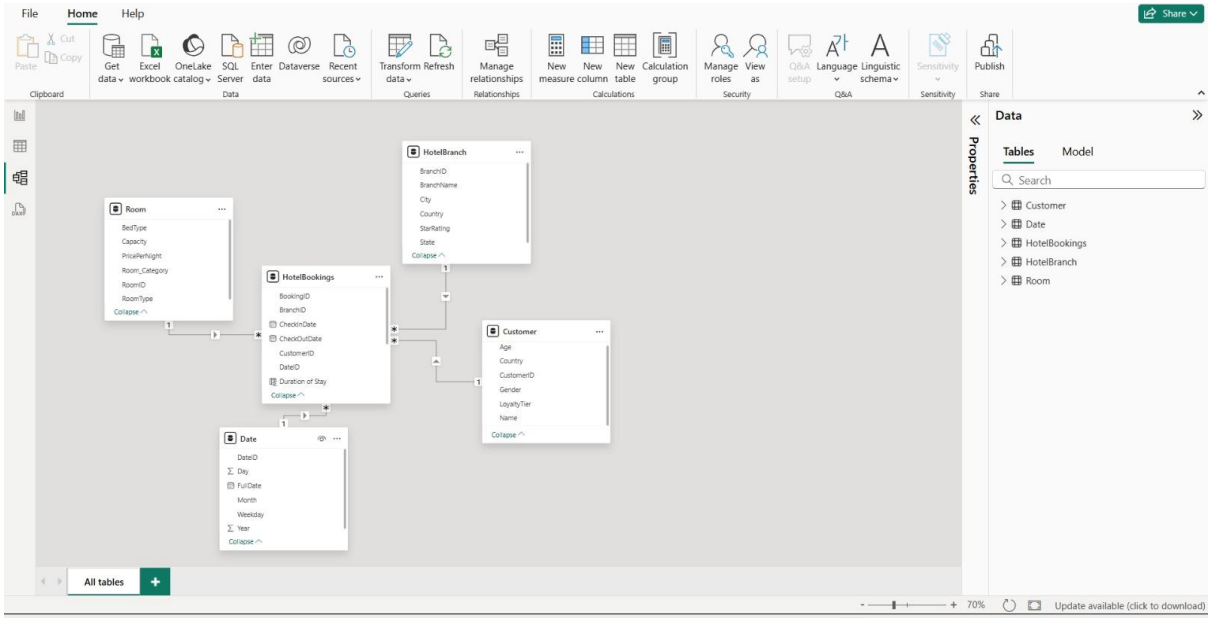
4. HotelBranch (Dimension Table)

- BranchID
- BranchName
- City
- State
- Country
- StarRating

5. Date (Dimension Table)

- DateID
- FullDate
- Day
- Month
- Year
- Weekday

Data Modeling



A star schema is implemented where:

- HotelBookings acts as the central fact table
- Customer, Room, HotelBranch, and Date act as dimension tables

This structure enables efficient filtering, aggregation, and cross-analysis across dimensions.

Data Transformation & Calculated Columns

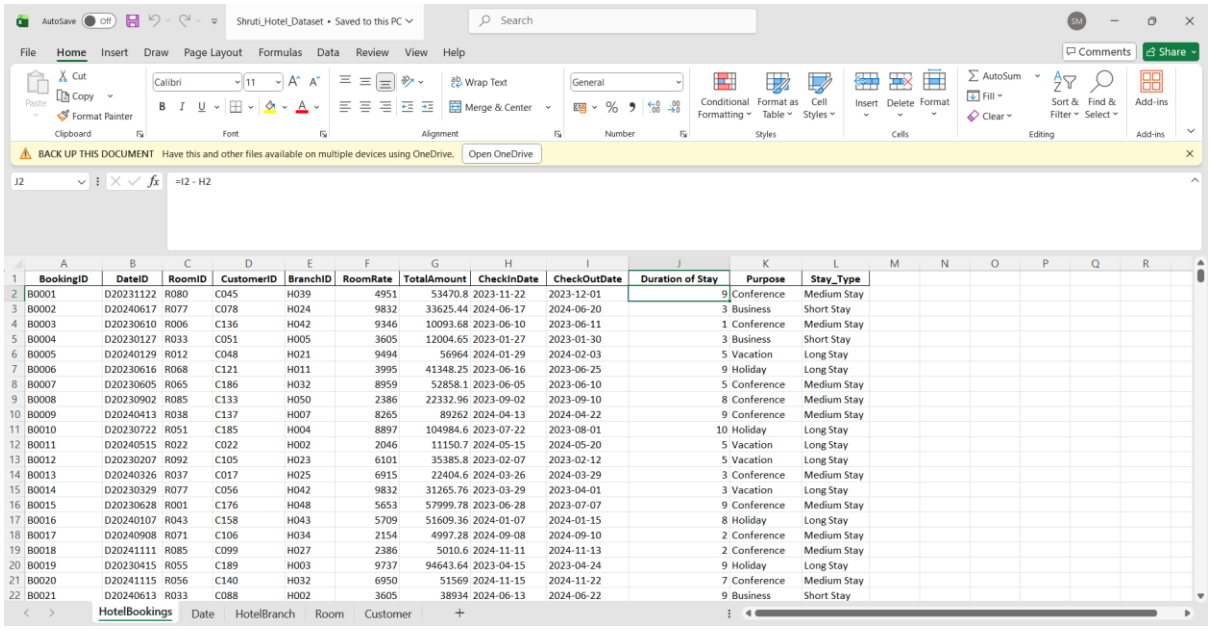
1. Duration of Stay

Calculated as the difference between check-out and check-in dates.

Excel Formula:

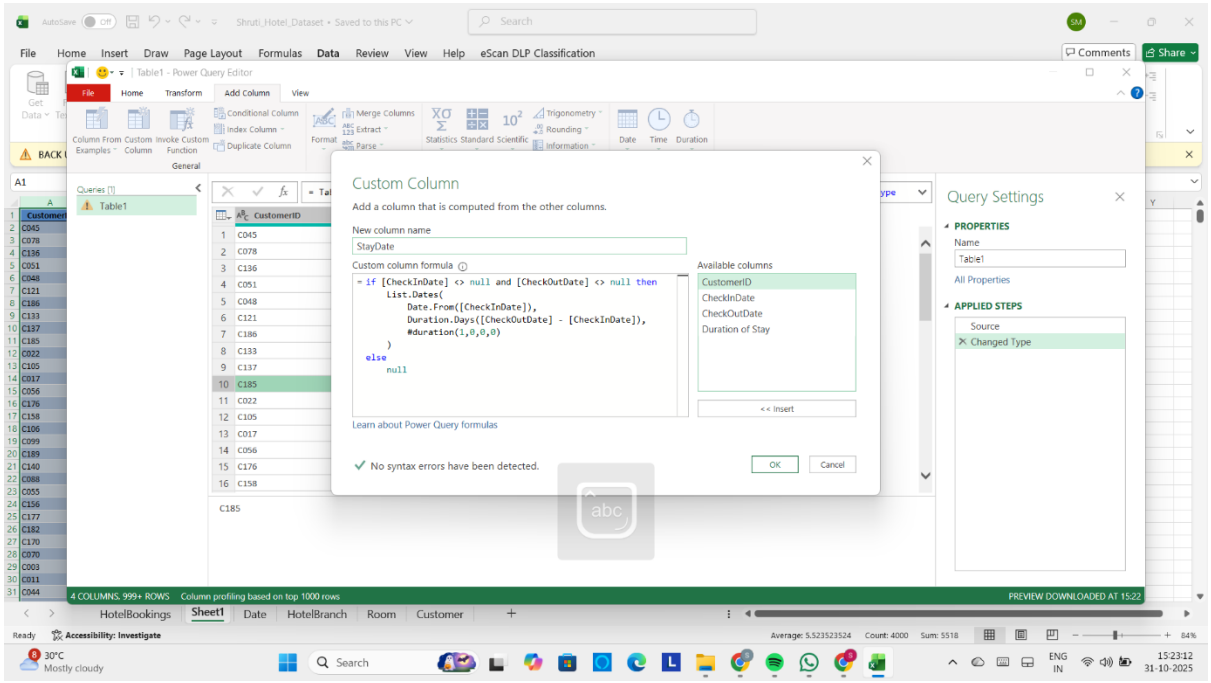
= CheckOutDate - CheckInDate

This helps analyze short, medium, and long guest stays.



BookingID	DateID	RoomID	CustomerID	BranchID	RoomRate	TotalAmount	CheckInDate	CheckOutDate	Duration of Stay	Purpose	Stay_Type
B0001	D20231122	R080	C045	H039	4951	53470.8	2023-11-22	2023-12-01	9	Conference	Medium Stay
B0002	D20240617	R077	C078	H024	9832	33625.44	2024-06-17	2024-06-20	3	Business	Short Stay
B0003	D20230610	R006	C136	H042	9346	10093.68	2023-06-10	2023-06-11	1	Conference	Medium Stay
B0004	D20230127	R033	C051	H005	3605	12004.65	2023-01-27	2023-01-30	3	Business	Short Stay
B0005	D20240129	R012	C048	H021	9494	56964	2024-01-29	2024-02-03	5	Vacation	Long Stay
B0006	D20230616	R068	C121	H011	3995	41348.25	2023-06-16	2023-06-25	9	Holiday	Long Stay
B0007	D20230605	R065	C186	H032	8959	52858.1	2023-06-05	2023-06-10	5	Conference	Medium Stay
B0008	D20230902	R085	C133	H050	2386	22332.96	2023-09-02	2023-09-10	8	Conference	Medium Stay
B0009	D20240413	R038	C137	H007	8265	89262	2024-04-13	2024-04-22	9	Conference	Medium Stay
B0010	D20230722	R051	C185	H004	8897	104984.6	2023-07-22	2023-08-01	10	Holiday	Long Stay
B0011	D20240515	R022	C022	H002	2046	11150.7	2024-05-15	2024-05-20	5	Vacation	Long Stay
B0012	D20230207	R092	C105	H023	6101	35385.8	2023-02-07	2023-02-12	5	Vacation	Long Stay
B0013	D20240326	R037	C017	H025	6915	22404.6	2024-03-26	2024-03-29	3	Conference	Medium Stay
B0014	D20230329	R077	C056	H042	9832	31265.76	2023-03-29	2023-04-01	3	Vacation	Long Stay
B0015	D20230628	R001	C176	H048	5653	57999.78	2023-06-28	2023-07-07	9	Conference	Medium Stay
B0016	D20240107	R043	C158	H043	5709	51609.36	2024-01-07	2024-01-15	8	Holiday	Long Stay
B0017	D20240908	R071	C106	H034	2154	4997.28	2024-09-08	2024-09-10	2	Conference	Medium Stay
B0018	D20241111	R085	C099	H027	2386	5010.6	2024-11-11	2024-11-13	2	Conference	Medium Stay
B0019	D20230415	R055	C189	H003	9737	94643.64	2023-04-15	2023-04-24	9	Holiday	Long Stay
B0020	D20241115	R056	C140	H032	6950	51569	2024-11-15	2024-11-22	7	Conference	Medium Stay
B0021	D20240613	R033	C088	H002	3605	38934	2024-06-13	2024-06-22	9	Business	Short Stay

2. StayDate



Each booking is expanded into individual stay dates using Power Query.

Power Query Logic:

- Generates a list of dates between CheckInDate and CheckOutDate
- Expands bookings into daily-level records

This enables daily occupancy and revenue analysis.

3. Stay Type Classification

Guests are categorized based on the purpose of visit.

Excel Formula:

=SWITCH(TRIM(UPPER(Purpose)),

"BUSINESS","Short Stay",

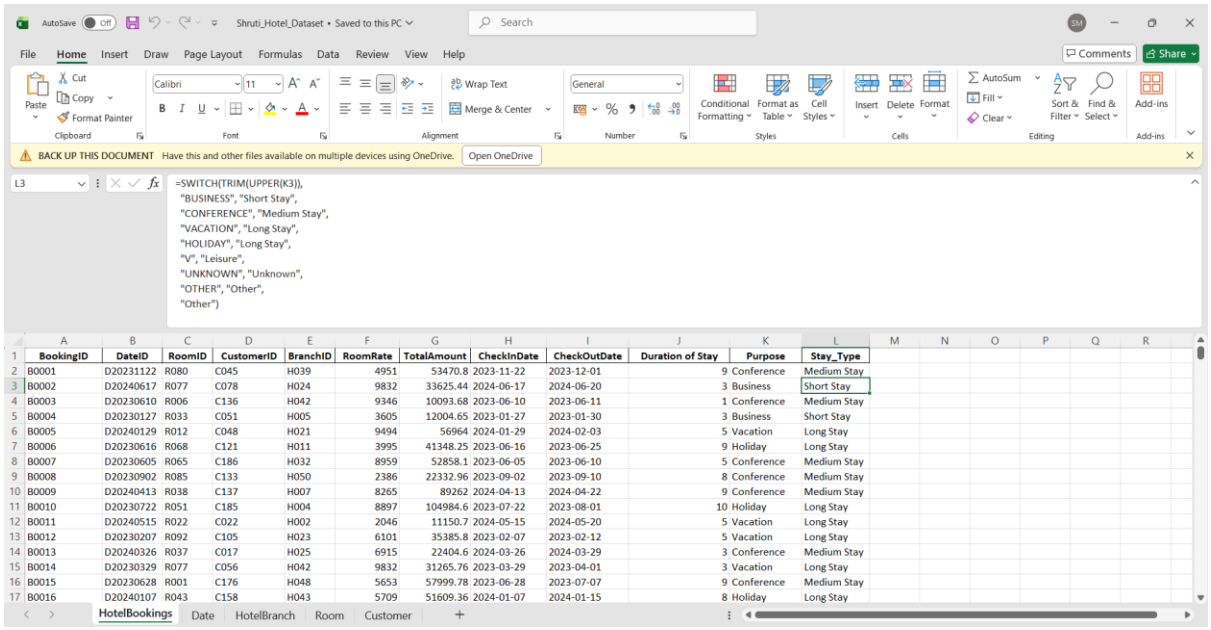
"CONFERENCE","Medium Stay",

"VACATION","Long Stay",

"HOLIDAY","Long Stay",

"OTHER","Other",

"Unknown")



BookingID	DateID	RoomID	CustomerID	BranchID	RoomRate	TotalAmount	CheckInDate	CheckOutDate	Duration of Stay	Purpose	Stay_Type
B0001	D20231122	R080	C045	H039	4951	53470.8	2023-11-22	2023-12-01	9	Conference	Medium Stay
B0002	D20240617	R077	C078	H024	9832	33625.44	2024-06-17	2024-06-20	3	Business	Short Stay
B0003	D20230610	R006	C136	H042	9346	10093.68	2023-06-10	2023-06-11	1	Conference	Medium Stay
B0004	D20230127	R033	C051	H005	3605	12004.65	2023-01-27	2023-01-30	3	Business	Short Stay
B0005	D20240129	R012	C048	H021	9494	56964	2024-01-29	2024-02-03	5	Vacation	Long Stay
B0006	D20230616	R068	C121	H011	3995	41348.25	2023-06-16	2023-06-25	9	Holiday	Long Stay
B0007	D20230605	R065	C186	H032	8959	52858.1	2023-06-05	2023-06-10	5	Conference	Medium Stay
B0008	D20230902	R085	C133	H050	2386	22332.96	2023-09-02	2023-09-10	8	Conference	Medium Stay
B0009	D20240413	R038	C137	H007	8265	89262	2024-04-13	2024-04-22	9	Conference	Medium Stay
B0010	D20230722	R051	C185	H004	8897	104984.6	2023-07-22	2023-08-01	10	Holiday	Long Stay
B0011	D20240515	R022	C022	H002	2046	11150.7	2024-05-15	2024-05-20	5	Vacation	Long Stay
B0012	D20230207	R092	C105	H023	6101	35385.8	2023-02-07	2023-02-12	5	Vacation	Long Stay
B0013	D20240326	R037	C017	H025	6915	22404.6	2024-03-26	2024-03-29	3	Conference	Medium Stay
B0014	D20230329	R077	C056	H042	9832	31265.76	2023-03-29	2023-04-01	3	Vacation	Long Stay
B0015	D20230628	R001	C176	H048	5653	57999.78	2023-06-28	2023-07-07	9	Conference	Medium Stay
B0016	D20240107	R043	C158	H043	5709	51609.36	2024-01-07	2024-01-15	8	Holiday	Long Stay

4. Room Category Mapping

Room types are grouped into pricing categories.

Excel Formula:

=SWITCH(TRIM(UPPER(RoomType)),

"SUITE","Premium",

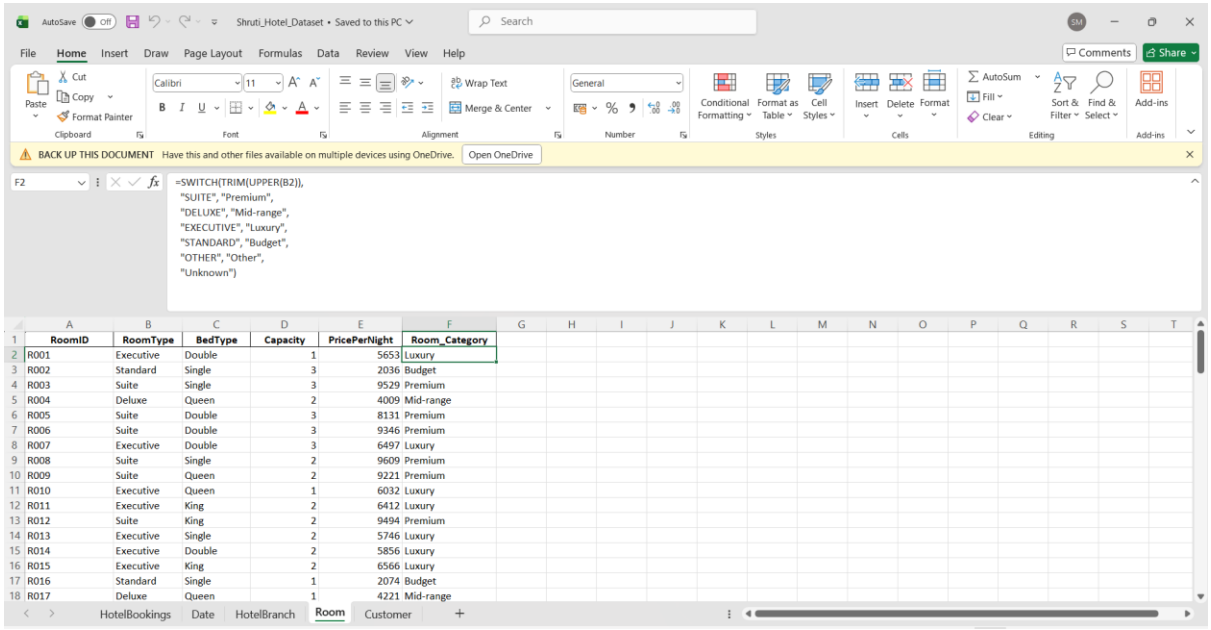
"DELUXE","Mid-range",

"EXECUTIVE","Luxury",

"STANDARD","Budget",

"OTHER","Other",

"Unknown")



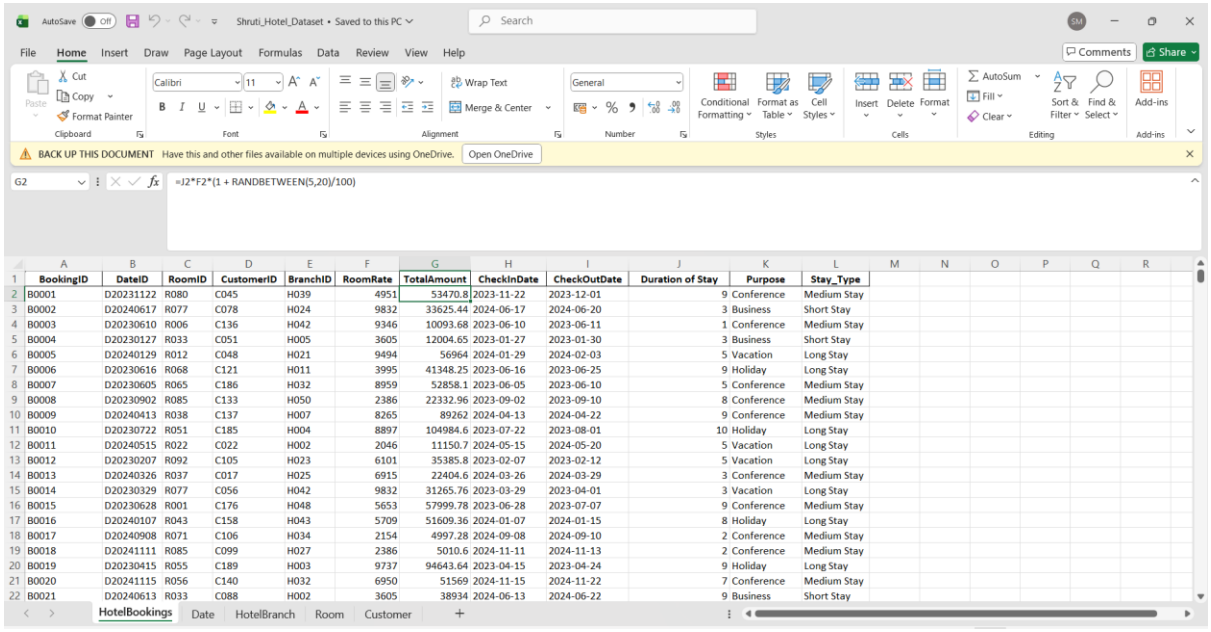
RoomID	RoomType	BedType	Capacity	PricePerNight	Room_Category
R001	Executive	Double	1	5653	Luxury
R002	Standard	Single	3	2036	Budget
R003	Suite	Single	3	9529	Premium
R004	Deluxe	Queen	2	4009	Mid-range
R005	Suite	Double	3	8131	Premium
R006	Suite	Double	3	9346	Premium
R007	Executive	Double	3	6497	Luxury
R008	Suite	Single	2	9609	Premium
R009	Suite	Queen	2	9221	Premium
R010	Executive	Queen	1	6032	Luxury
R011	Executive	King	2	6412	Luxury
R012	Suite	King	2	9494	Premium
R013	Executive	Single	2	5746	Luxury
R014	Executive	Double	2	5856	Luxury
R015	Executive	King	2	6566	Luxury
R016	Standard	Single	1	2074	Budget
R017	Deluxe	Queen	1	4221	Mid-range

5. Total Amount Calculation

Room revenue is calculated with a controlled variation to simulate real pricing behavior.

Excel Formula:

= RoomRate * DurationOfStay * (1 + RANDBETWEEN(5,20)/100)



BookingID	DateID	RoomID	CustomerID	BranchID	RoomRate	TotalAmount	CheckInDate	CheckOutDate	Duration of Stay	Purpose	Stay_Type
B0001	D20231122	R080	C045	H039	4951	53470.8	2023-11-22	2023-12-01	9	Conference	Medium Stay
B0002	D20240617	R077	C078	H024	9832	33625.44	2024-06-17	2024-06-20	3	Business	Short Stay
B0003	D20230610	R006	C136	H042	9346	10093.68	2023-06-10	2023-06-11	1	Conference	Medium Stay
B0004	D20230127	R033	C051	H005	3605	12004.65	2023-01-27	2023-01-30	3	Business	Short Stay
B0005	D20240129	R012	C048	H021	9494	56964	2024-01-29	2024-02-03	5	Vacation	Long Stay
B0006	D20230616	R068	C121	H011	3995	41348.25	2023-06-16	2023-06-25	9	Holiday	Long Stay
B0007	D20230605	R065	C186	H032	8959	52858.1	2023-06-05	2023-06-10	5	Conference	Medium Stay
B0008	D20230902	R085	C133	H050	2386	22332.96	2023-09-02	2023-09-10	8	Conference	Medium Stay
B0009	D20240413	R038	C137	H007	8265	89262	2024-04-13	2024-04-22	9	Conference	Medium Stay
B0010	D20230722	R051	C185	H004	8897	104984.6	2023-07-22	2023-08-01	10	Holiday	Long Stay
B0011	D20240515	R022	C022	H002	2046	11150.7	2024-05-15	2024-05-20	5	Vacation	Long Stay
B0012	D20230207	R092	C105	H023	6101	35385.8	2023-02-07	2023-02-12	5	Vacation	Long Stay
B0013	D20240326	R037	C017	H025	6915	22404.6	2024-03-26	2024-03-29	3	Conference	Medium Stay
B0014	D20230329	R077	C056	H042	9832	31265.76	2023-03-29	2023-04-01	3	Vacation	Long Stay
B0015	D20230628	R001	C176	H048	5653	57999.78	2023-06-28	2023-07-07	9	Conference	Medium Stay
B0016	D20240107	R043	C158	H043	5709	51609.36	2024-01-07	2024-01-15	8	Holiday	Long Stay
B0017	D20240908	R071	C106	H034	2154	4997.28	2024-09-08	2024-09-10	2	Conference	Medium Stay
B0018	D20241111	R085	C099	H027	2386	5010.6	2024-11-11	2024-11-13	2	Conference	Medium Stay
B0019	D20230415	R055	C189	H003	9737	94643.64	2023-04-15	2023-04-24	9	Holiday	Long Stay
B0020	D20241115	R056	C140	H032	6950	51569	2024-11-15	2024-11-22	7	Conference	Medium Stay
B0021	D20240613	R033	C088	H002	3605	38934	2024-06-13	2024-06-22	9	Business	Short Stay

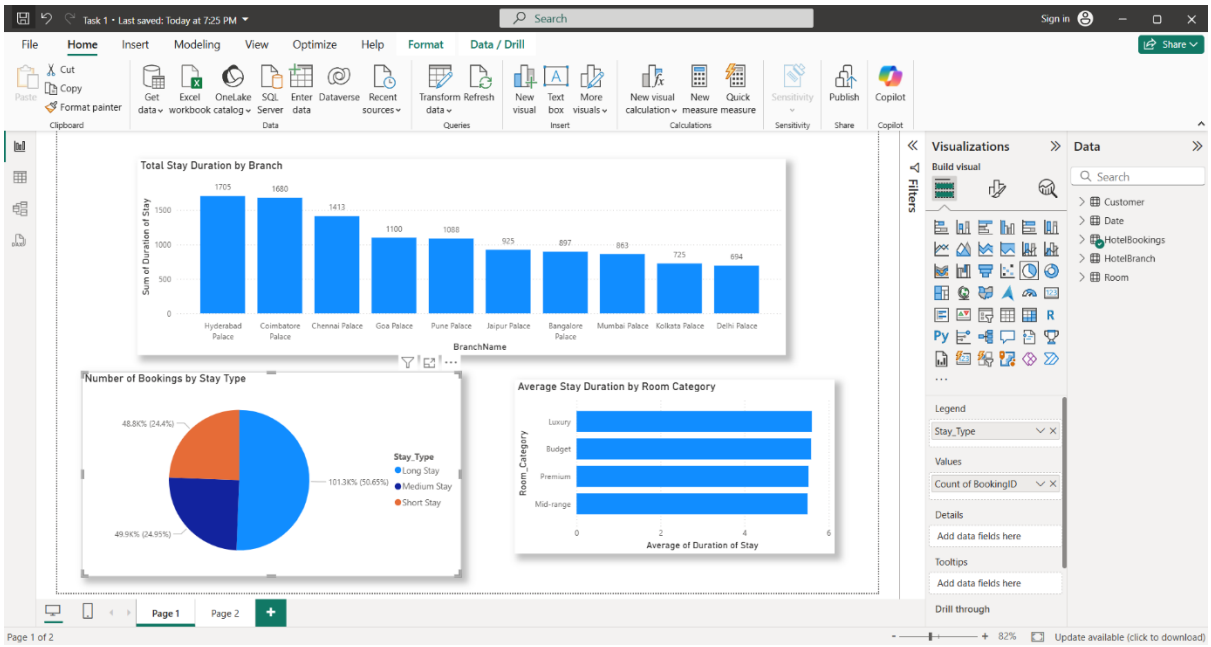
6. Room Booked Aggregation

Daily room bookings are calculated using **Group By** in Power Query on:

- StayDate
- RoomID

This supports occupancy calculations in later modules.

Visualizations & Analysis (Output)



1. Total Stay Duration by Branch

Visual: Bar Chart
Insight:
Shows which hotel branches generate longer guest stays.
Branches like Hyderabad Palace and Coimbatore Palace show higher total stay durations, indicating stronger guest retention.

2. Number of Bookings by Stay Type

Visual: Pie Chart
Insight:
Displays distribution of:

- Short Stay
- Medium Stay

- Long Stay

Helps understand dominant guest behavior patterns.

3. Average Stay Duration by Room Category

Visual: Horizontal Bar Chart
Insight:
Luxury and Premium rooms show longer average stays, indicating higher-value guests prefer premium categories.

Key Insights from Module 1

- Long-stay bookings contribute significantly to total occupancy.
- Premium and Luxury rooms attract guests with longer durations.
- Certain branches consistently outperform others in guest retention.
- Daily-level data preparation enables accurate occupancy and revenue tracking in later modules.

Conclusion

Module 1 establishes a robust analytical base by transforming raw booking data into structured, enriched datasets.
By analyzing stay duration, room categories, and branch performance, this module provides essential insights into guest behavior.

The clean data model and calculated fields created in this module serve as the foundation for advanced analytics in Occupancy Metrics, Revenue Analysis, Forecasting, and Revenue Strategy Dashboards in subsequent modules.