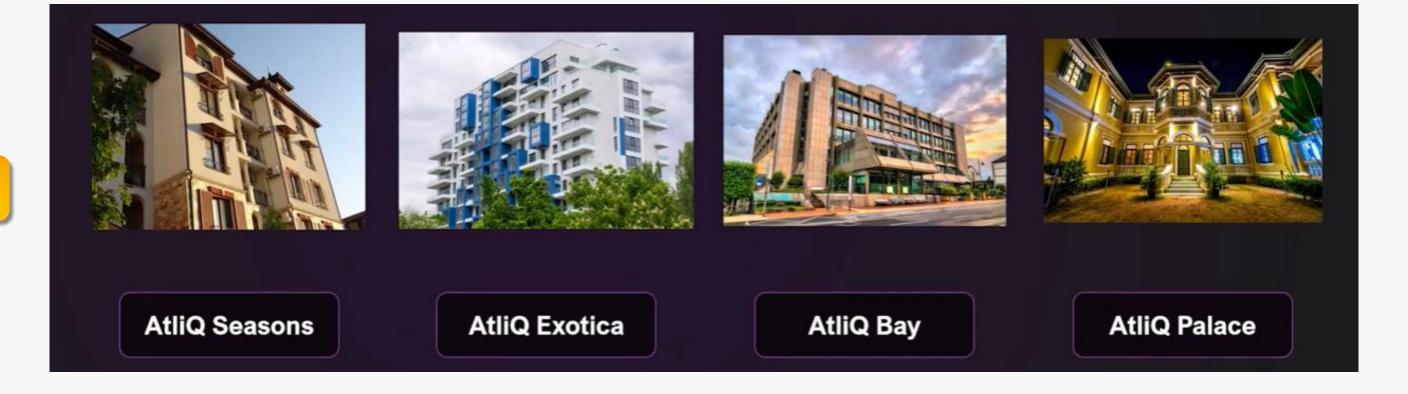
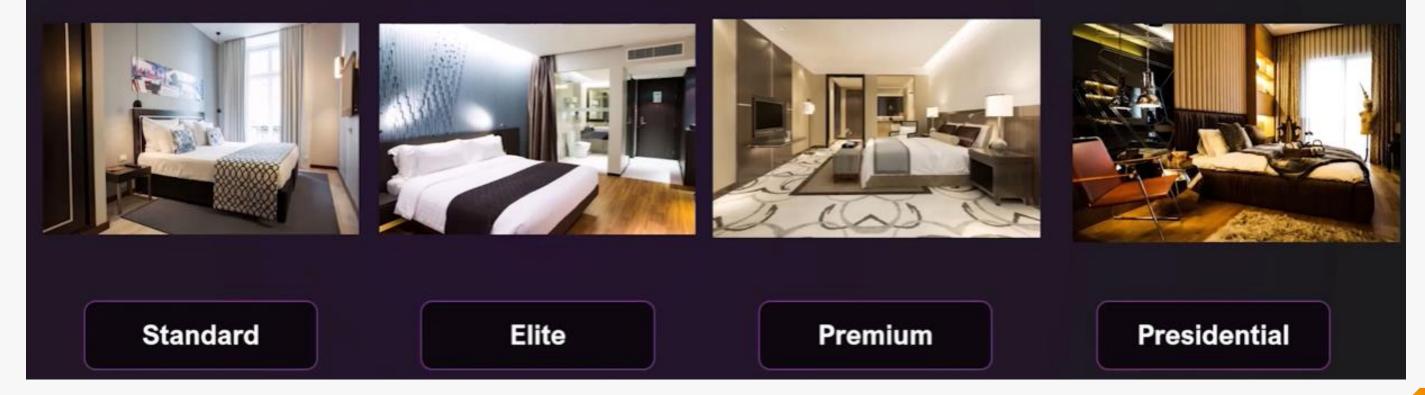


**Properties** 

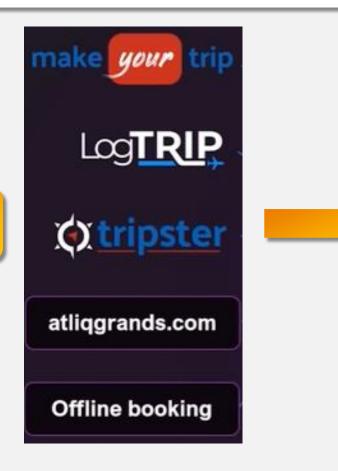


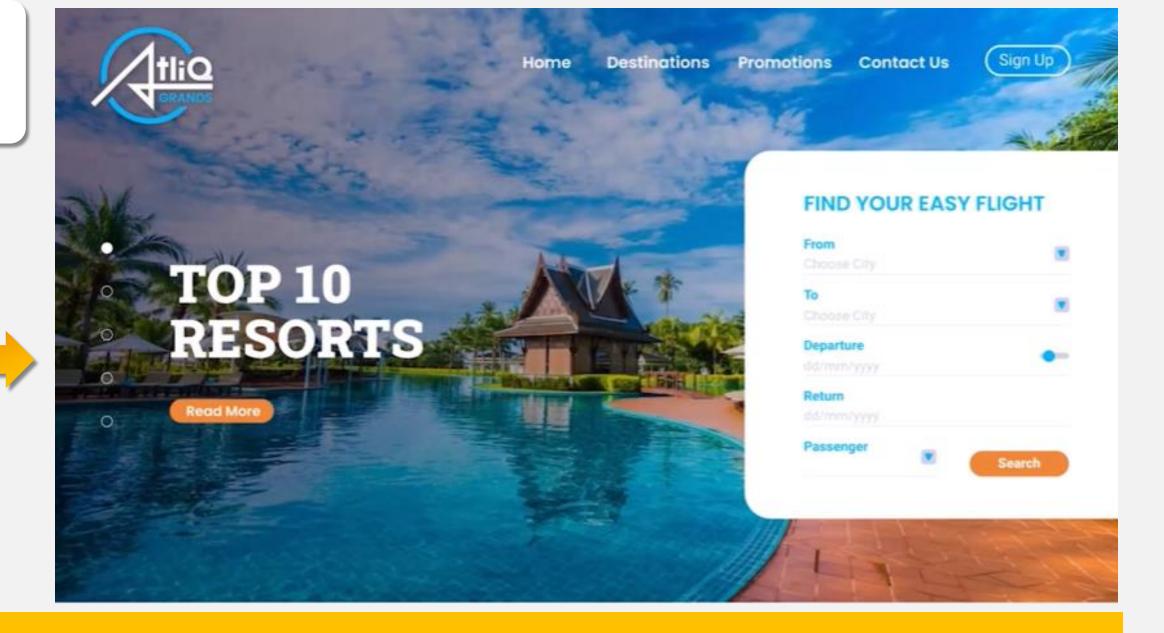
**Room Class** 





#### **Understanding Business Problem**





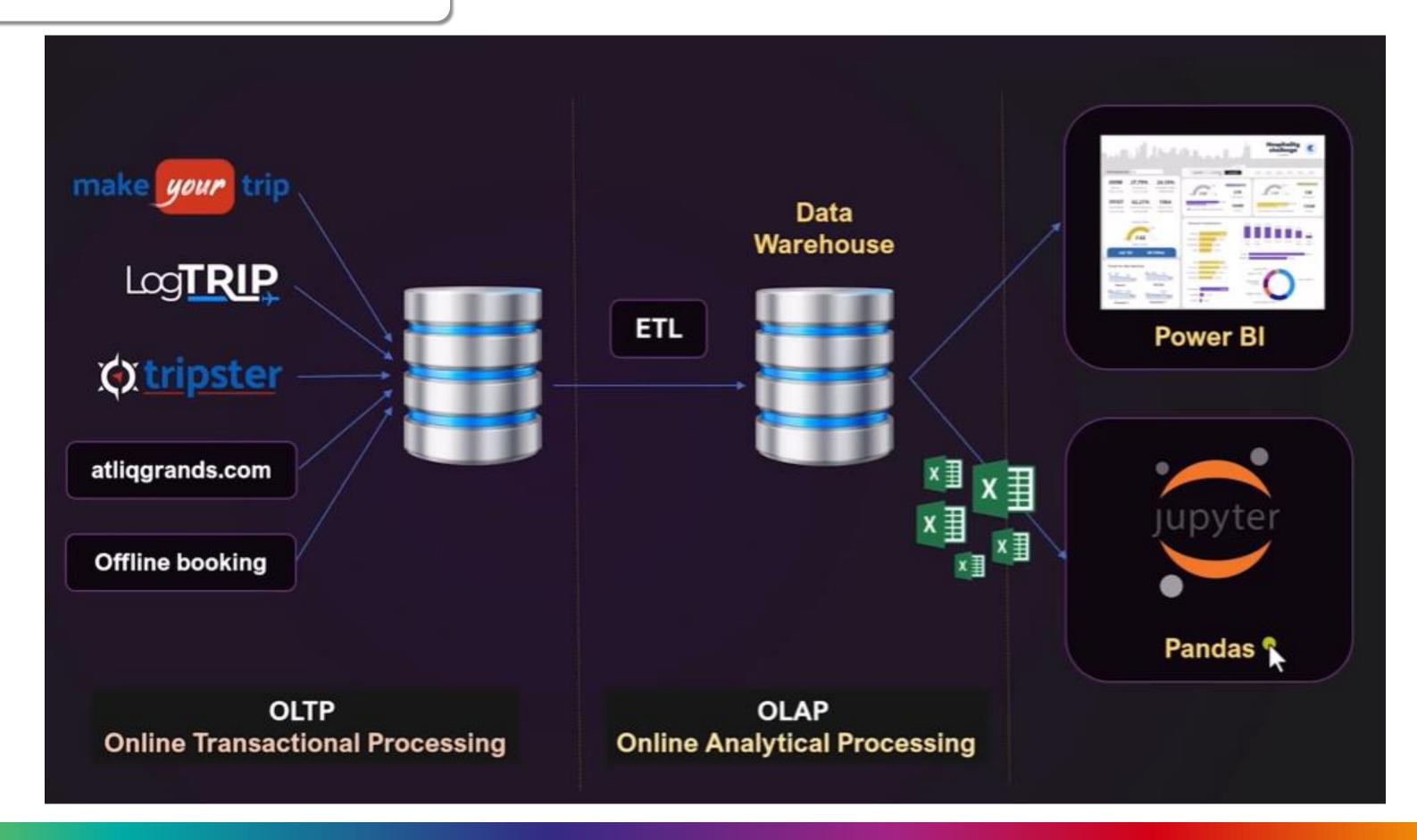


**Booking Platforms** 

### Problem Statement:

- > AtliQ Grands, a renowned luxury hotel chain in India, is facing challenges in market share and revenue due to competition and management decisions. To counter this, the managing director seeks to leverage Business and Data Intelligence but lacks an in-house analytics team. As a data analyst, my task is to utilize sample data and a mock-up dashboard to:
- Create specified metrics.
- Develop a dashboard as per stakeholder requirements.
- Generate additional insights beyond the given metrics and mock-up.







Data Analyst Data Engineer Usawa Landon Data Analysis pandas

#### **Data Analyst**



**Data Engineer** 







**Data Analysis** 





245]: import pandas as pd

==> 1. Data Import and Data Exploration

#### **Datasets**

We have 5 csv file

- dim\_date.csv
- dim\_hotels.csv
- dim\_rooms.csv
- fact\_aggregated\_bookings
- fact\_bookings.csv

#### **Loading CSV Files in Jupyter Notebook**



```
df_date = pd.read_csv('datasets/dim_date.csv')
df_hotels = pd.read_csv('datasets/dim_hotels.csv')
df_rooms = pd.read_csv('datasets/dim_rooms.csv')
df_agg_bookings = pd.read_csv('datasets/fact_aggregated_bookings.csv')

df_bookings = pd.read_csv('datasets/fact_bookings.csv')
```



(134590, 12)

#### **Exploratory Data Analysis**

#### ==> 2. Data Cleaning

df\_bookings.describe()

	property_id	no_guests	ratings_given	revenue_generated	revenue_realized
count	134590.000000	134587.000000	56683.000000	1.345900e+05	134590.000000
mean	18061.113493	2.036170	3.619004	1.537805e+04	12696.123256
std	1093.055847	1.034885	1.235009	9.303604e+04	6928.108124
min	16558.000000	-17.000000	1.000000	6.500000e+03	2600.000000
25%	17558.000000	1.000000	3.000000	9.900000e+03	7600.000000
50%	17564.000000	2.000000	4.000000	1.350000e+04	11700.000000
<b>75</b> %	18563.000000	2.000000	5.000000	1.800000e+04	15300.000000
max	19563.000000	6.000000	5.000000	2.856000e+07	45220.000000
df_booki	ngs.shape				

- After running "describe" function, the values in "no of Guests" column cannot be negative, this is an error, need to check how many more negative values are there and need to clean up.
- ☐ To find out, how many rows & columns are there run "shape" function



#### **Exploratory Data Analysis**

#### (1) Clean invalid guests

df\_bookings[df\_bookings.no\_guests<=0]</pre>

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1
3	May012216558RT14	16558	28-04-22	1/5/2022	2/5/2022	-2.0	RT1
17924	May122218559RT44	18559	12/5/2022	12/5/2022	14-05-22	-10.0	RT4
18020	May122218561RT22	18561	8/5/2022	12/5/2022	14-05-22	-12.0	RT2
18119	May122218562RT311	18562	5/5/2022	12/5/2022	17-05-22	-6.0	RT3
18121	May122218562RT313	18562	10/5/2022	12/5/2022	17-05-22	-4.0	RT3
56715	Jun082218562RT12	18562	5/6/2022	8/6/2022	13-06-22	-17.0	RT1
119765	Jul202219560RT220	19560	19-07-22	20-07-22	22-07-22	-1.0	RT2
134586	Jul312217564RT47	17564	30-07-22	31-07-22	1/8/2022	-4.0	RT4

As you can see above, number of guests having less than zero value represents data error. We can ignore these records.

df\_bookings = df\_bookings[df\_bookings.no\_guests>0]

 ${\tt df\_bookings.shape}$ 

(134578, 12)

Check how many negative values are available in df\_bookings data frame.
12 records are having negative values, and it is a data error, and it can be

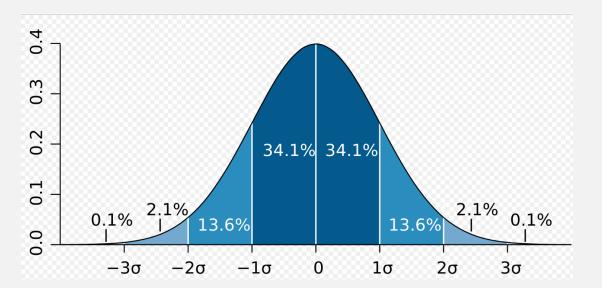
removed.

→ After removing the negative values now, the data frame has "134578" rows, This is how you can cleanup unwanted data and create the errorless data frame.



#### **Exploratory Data Analysis**

```
(2) Outlier removal in revenue generated
df_bookings.revenue_generated.min(), df_bookings.revenue_generated.max()
       28560000)
(6500,
df bookings.revenue generated.mean(), df bookings.revenue generated.median()
(15378.036937686695, 13500.0)
avg, std = df_bookings.revenue_generated.mean(), df_bookings.revenue_generated.std()
avg, std
(15378.036937686695, 93040.1549314641)
higher_limit = avg + 3*std
higher_limit
294498.50173207896
lower_limit = avg - 3*std
lower limit
-263742.4278567056
df_bookings[df_bookings.revenue_generated<=0]</pre>
  booking_id property_id booking_date check_in_date checkout_date no_guests room_category
```



- Outlier removal by using 3 standard deviation technique. If value is greater than 3 standard deviation, then it can be considered as an Outlier, Here Higher limit is "2,94,498", so definitely "
  "2,85,60,000" is an outlier and it has to be removed in revenue generated column.
- →□ We also know that revenue generated cannot be negative, so there are now negative records.

### **Exploratory Data Analysis**

[275]:	df_bookings[df_bookings.revenue_generated>higher_limit]								⊕ ↑	
[275]:	perty_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given	booking_status	revenue_generated
	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip	5.0	Checked Out	9100000
	16559	29-04-22	1/5/2022	2/5/2022	6.0	RT3	direct online	NaN	Checked Out	28560000
	16562	28-04-22	1/5/2022	4/5/2022	2.0	RT2	direct offline	3.0	Checked Out	12600000
	17559	26-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN	Cancelled	2000000
	16562	21-07-22	28-07-22	29-07-22	2.0	RT2	direct online	3.0	Checked Out	10000000
	4									
[276]:	<pre>df_bookings = df_bookings[df_bookings.revenue_generated&lt;=higher_limit] df_bookings.shape</pre>									
[276]:	(134573)	12)								

- ☐ 5 rows with greater than higher limit was available in the revenue generated column, and it was removed by modifying the data frame.
- ☐ After removing the data errors now the df\_bookings is having "134573" rows.





#### **Data Transformation & Modelling**

#### ==> 3. Data Transformation ¶

Create occupancy percentage column

df\_agg\_bookings.head(3)

	property_id	check_in_date	room_category	successful_bookings	capacity
0	16559	1-May-22	RT1	25	30.0
1	19562	1-May-22	RT1	28	30.0
2	19563	1-May-22	RT1	23	30.0

df\_agg\_bookings['occ\_pct'] = df\_agg\_bookings.apply(lambda row: row['successful\_bookings']/row['capacity'], axis=1)

You can use following approach to get rid of SettingWithCopyWarning

new\_col = df\_agg\_bookings.apply(lambda row: row['successful\_bookings']/row['capacity'], axis=1)
df\_agg\_bookings = df\_agg\_bookings.assign(occ\_pct=new\_col.values)
df\_agg\_bookings.head(3)

	property_id	check_in_date	room_category	$successful\_bookings$	capacity	occ_pct
0	16559	1-May-22	RT1	25	30.0	0.833333
1	19562	1-May-22	RT1	28	30.0	0.933333
2	19563	1-May-22	RT1	23	30.0	0.766667

Convert it to a percentage value

df\_agg\_bookings['occ\_pct'] = df\_agg\_bookings['occ\_pct'].apply(lambda x: round(x\*100, 2))
df\_agg\_bookings.head(3)

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct
0	16559	1-May-22	RT1	25	30.0	83.33
1	19562	1-May-22	RT1	28	30.0	93.33
2	19563	1-May-22	RT1	23	30.0	76.67
						4

There are various types of data transformations that you may have to perform based on the need.

Few examples of data transformations are,

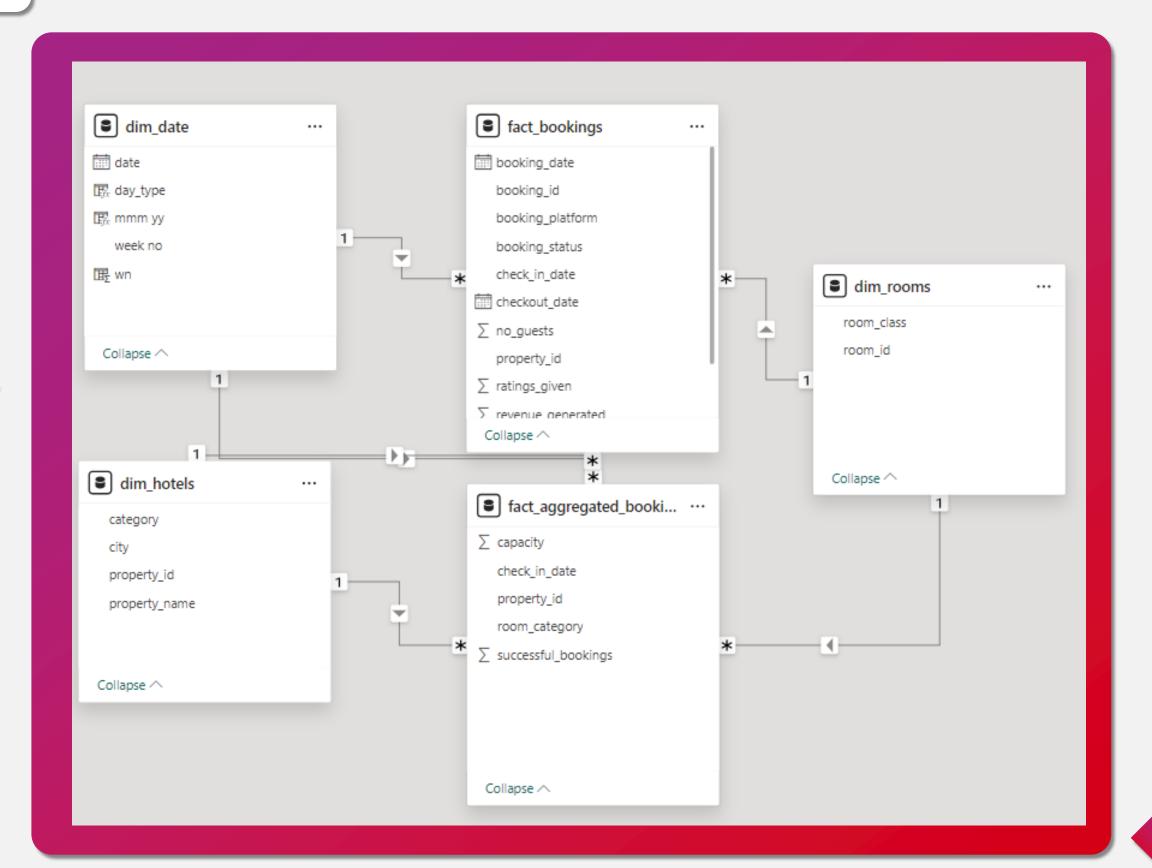
- Creating new columns
- Normalization
- Merging data
- Aggregation



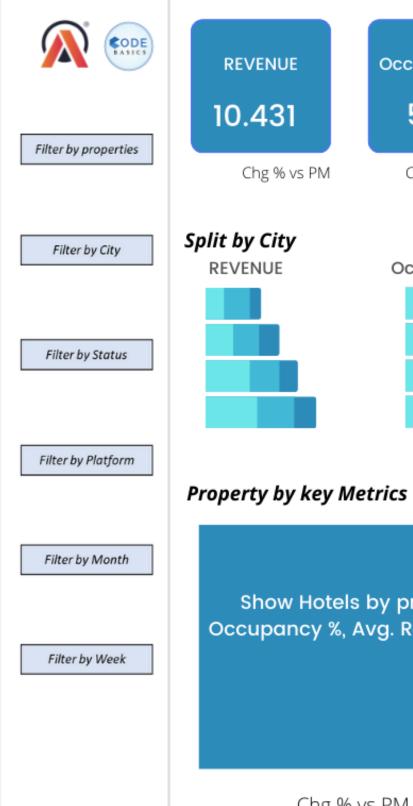


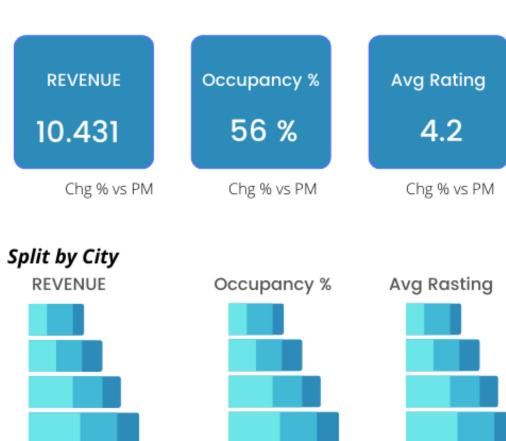
### **Data Transformation & Modelling**

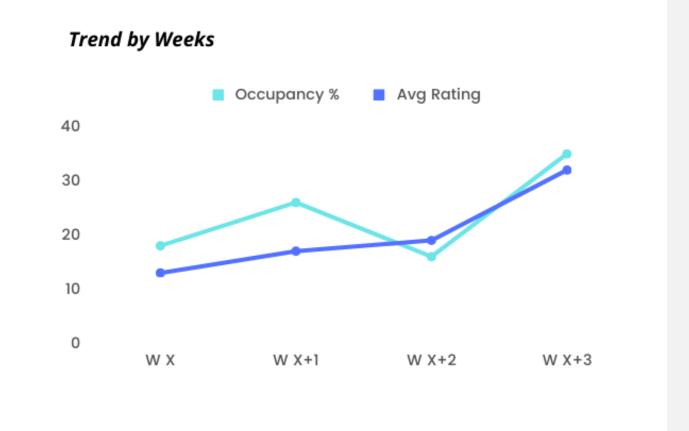
**Star Schema Data Modelling** 



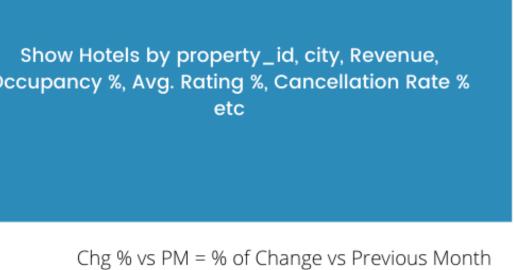
### **Mock-up Dashboard**

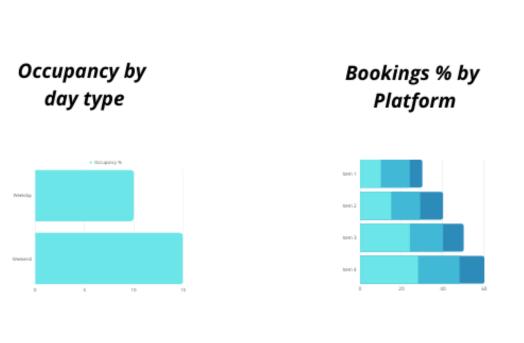






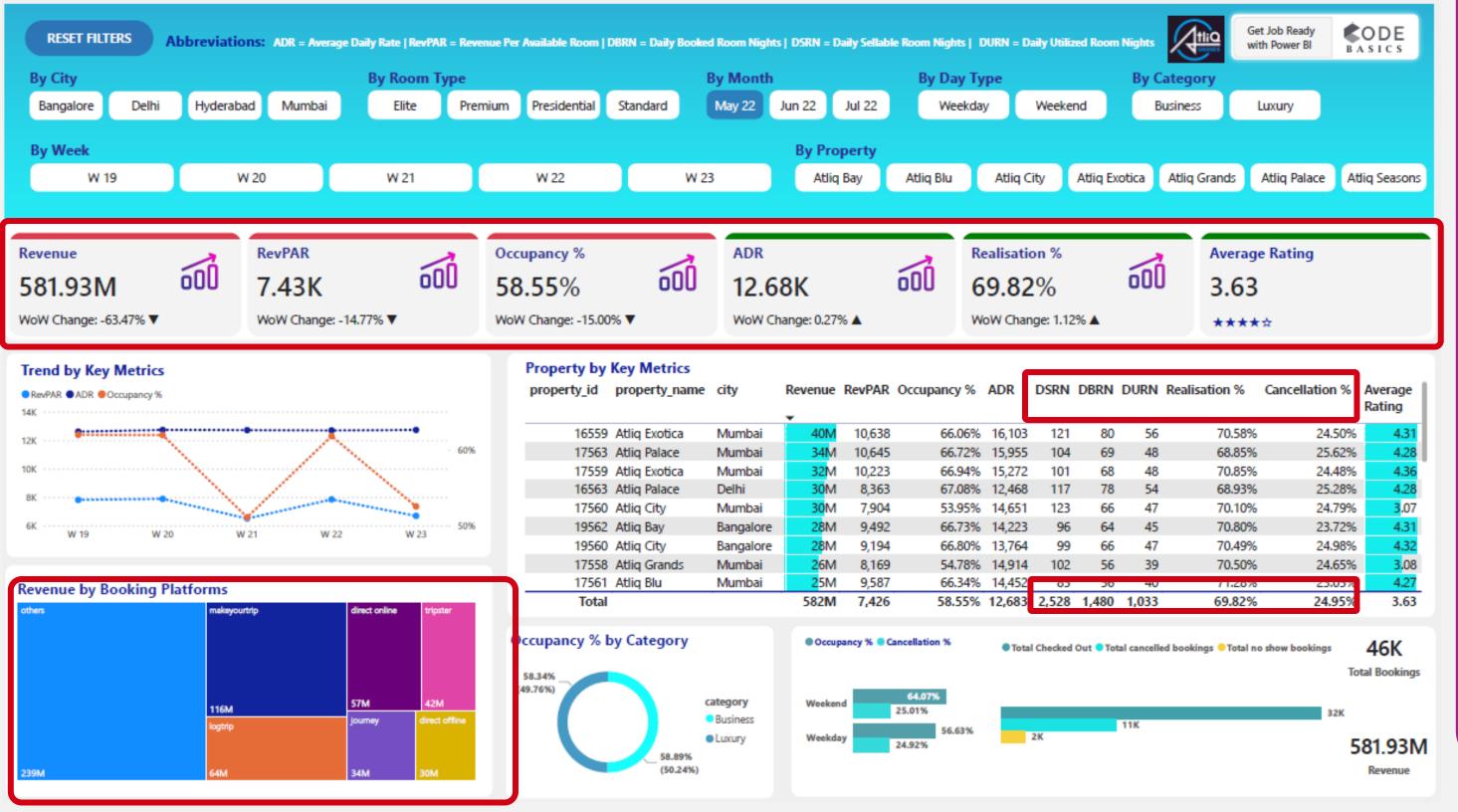






#### **Power BI – Hospitality Main View**

#### May 22 - Statistics



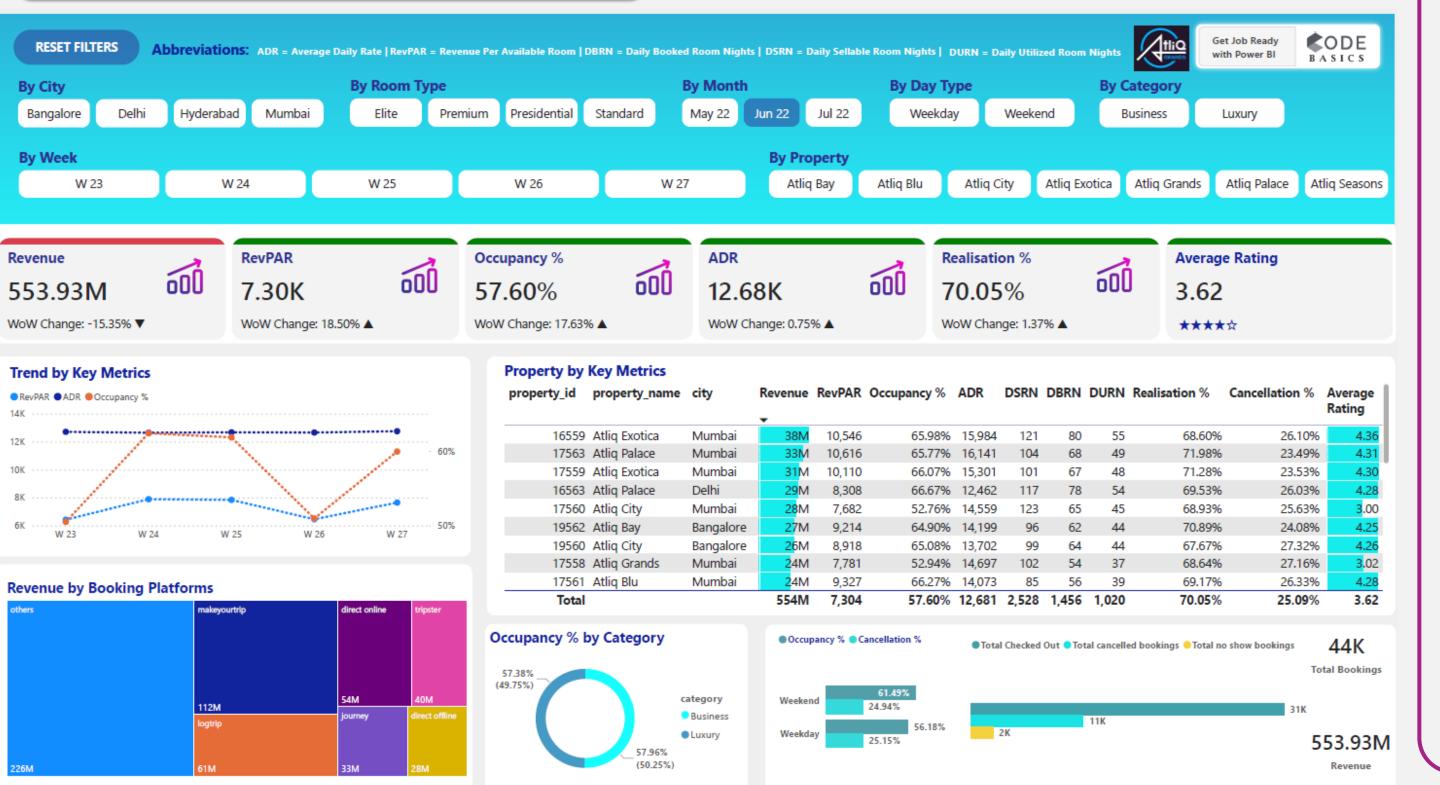
"May 2022" month data analysis are mentioned below.

- Revenue generated is **582 M**.
- Average Revenue per available room is 7.43K.
- Occupancy % is very low at around
   58.55% need to be improved
- Average daily rate is 12.68K is constant it should be changed on Weekends to generate more revenue
- Realisation % is **69.82%**
- Average Rating is only 3.63 need to be improved.
- DSRN is **2,528**.
- DBRN is **1,480**.
- DURN is only 1,033, so daily utilized rooms are very less because of less Occupancy % and cancellation % is also around 25 %, it must be reduced
- Make your trip is one of the booking platform which is contributing around 19-20% to the revenue.
- However major revenue around 40% is coming from unrecognized sources which needs to identified for better marketing efforts.
- Naturally on Weekend's the Occupancy % is 64.07% compared to Weekday's i.e., 56.63%



#### **Power BI – Hospitality Main View**

#### Jun 22 - Statistics

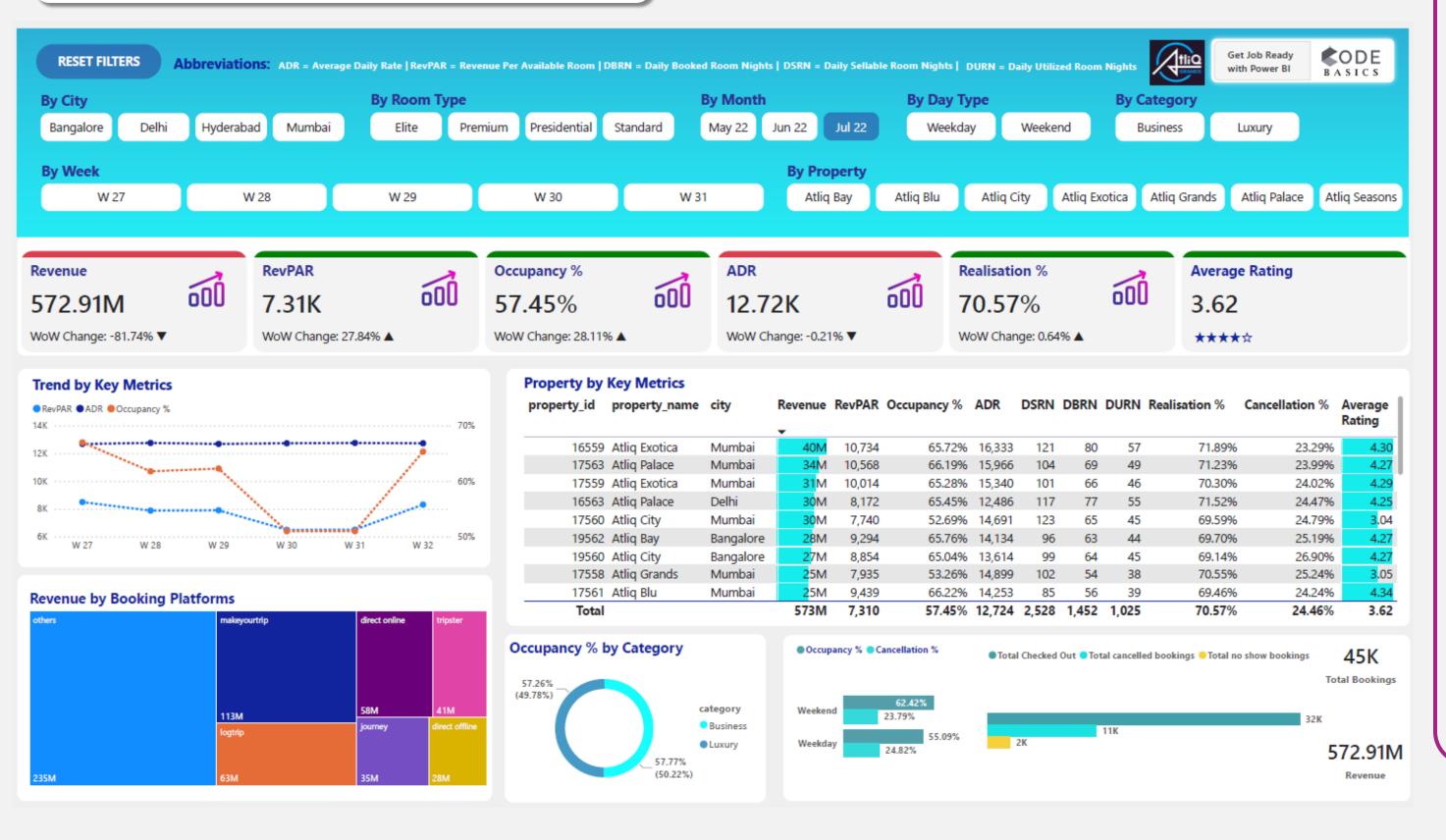


- "Jun 2022" month data analysis are mentioned below.
- Revenue generated is **554 M**.
- Average Revenue per available room is
   7.30K
- Occupancy % is very low at around
   57.60% need to be improved
- Average daily rate is 12.68K is constant it should be changed on Weekends to generate more revenue
- Realisation % is **70.05%**
- Average Rating is reduced to 3.62 compared to May month. need to be improved.
- DSRN is **2,528**.
- DBRN is **1,480.**
- DURN is only 1,020 again reduced, so daily utilized rooms are very less because of less Occupancy % and cancellation % is also around 25 %, it must be reduced
- Make your trip is one of the booking platform which is contributing around 19-20% to the revenue.
- However major revenue around 40% is coming from unrecognized sources which needs to identified for better marketing efforts.
- Naturally on Weekend's the Occupancy % is 61.49% compared to Weekday's i.e., 56.18%.



#### **Power BI – Hospitality Main View**

#### Jul 22 - Statistics



"Jul 2022" month data analysis are mentioned below.

- Revenue generated is **573 M**.
- Average Revenue per available room is 7.31K
- Occupancy % is very low at around
   57.45% need to be improved
- Average daily rate is 12.72K is constant it should be changed on Weekends to generate more revenue
- Realisation % is 70.57%
- Average Rating is reduced to 3.62 compared to May month. need to be improved.
- DSRN is **2,528**.
- DBRN is 1,452.
- DURN is only 1,025 again reduced, so daily utilized rooms are very less because of less Occupancy % and cancellation % is also around 25 %, it must be reduced
- Make your trip is one of the booking platform which is contributing around 19-20% to the revenue.
- However major revenue around 40% is coming from unrecognized sources which needs to identified for better marketing efforts.
- Naturally on Weekend's the Occupancy % is 62.42% compared to Weekday's i.e., 55.09%.



### 8 Performance Analytics

● Elite ● Premium ● Presidential ● Standard

#### **City & Room Class Analytics**

● Elite ● Premium ● Presidential ● Standard



● Elite ● Premium ● Presidential ● Standard

#### City Wise Analytics:

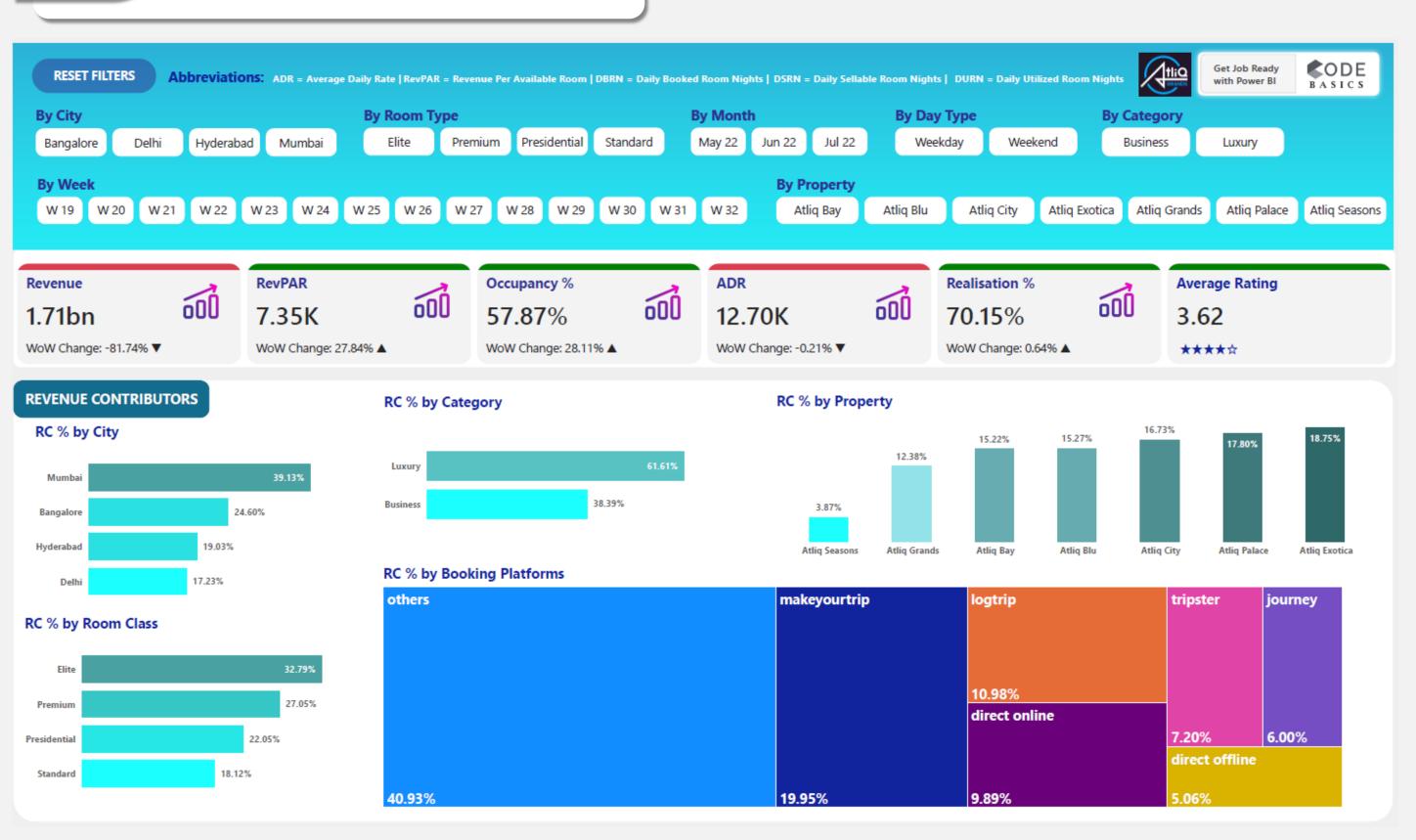
- Revenue: In revenue generation
   Mumbai is in 1<sup>st</sup> place irrespective of all the months and 2<sup>nd</sup> place goes to Bangalore & Last place goes to Delhi.
- Occupancy %: Delhi is leading in 1<sup>st</sup> position and Bangalore is in last position.
- Cancellation %: Delhi is having more cancellation% of 25.63% compared to other cities

#### **Room Class Analytics:**

- Revenue: Elite is in the first place in contributing the revenue and the last place goes to standard room class.
- Occupancy %: Presidential Suite is having highest occupancy % compared to others.
- Cancellation %: Elite is having more cancellation% of 25.61% compared to other room types.



# **9** Revenue Contributors



#### Revenue Contributors:

- RC % by City: Mumbai is in 1<sup>st</sup> place with RC % of 39.13% by generating revenue irrespective of all the months and 2<sup>nd</sup> place goes to Bangalore & Last place goes to Delhi.
- RC % by Room Type: Elite is in 1st place with RC % of 32.73%
- RC % by Category: Luxury Category
  is in first place by generating 61.61%
  of revenue compared to Business
  category.
- RC % by Booking Platforms:
  - Make your trip is one of the booking platform which is contributing around 19-20% to the revenue.
  - However major revenue around 40% is coming from unrecognized sources which needs to identified for better marketing efforts.
- RC % by Property:
  - AtliQ Exotica is in 1st place with RC % of 18.75% and 2<sup>nd</sup> place goes to AtliQ Palace with RC% of 17.80%
  - AtliQ Seasons is the worst performing property with rating 2.29 & RC % is 3.87%



## 10 Insights Summary

#### **Revenue Insights Summary:**

- Month-wise Revenue:
  - May 2022 recorded the highest revenue of 582M, followed by June 2022 with 554M and July 2022 with 573M.
- Customer Segment Analysis:
  - RC % by City:
    - Mumbai leads with an RC % of 39.13%, followed closely by Bangalore, while Delhi trails behind.
  - RC % by Room Type:
    - Elite rooms dominate with a significant RC % of 32.73%.
  - RC % by Category:
    - Luxury category outperforms, contributing 61.61% of revenue compared to the Business category.
- Marketing Channel Analysis:
  - RC % by Booking Platforms:
    - "Make Your Trip" accounts for approximately 19-20% of revenue.
    - Approximately 40% of revenue originates from unidentified sources, warranting focused identification efforts for enhanced marketing strategies.
- Property Performance Analysis:
  - RC % by Property:
    - AtliQ Exotica leads with an RC % of 18.75%, followed closely by AtliQ Palace at 17.80%.
    - AtliQ Seasons lags significantly with a rating of 2.29 and an RC % of 3.87%, indicating room for improvement.



### BECOME AN AI/DATA PROFESSIONAL



FROM UPSKILLING TO JOB READY



CODEBASICS.10



Mr. DHAVAL PATEL





Mr. HEMANAND VADIVEL