

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_csv('/content/fact_bookings.csv')
```

```
data.head()
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests
0	May012216558RT11	16558	27-04-2022	01-05-2022	02-05-2022	3
1	May012216558RT12	16558	30-04-2022	01-05-2022	02-05-2022	2
2	May012216558RT13	16558	28-04-2022	01-05-2022	04-05-2022	2
3	May012216558RT14	16558	28-04-2022	01-05-2022	02-05-2022	2
4	May012216558RT15	16558	27-04-2022	01-05-2022	02-05-2022	4

```
data.shape
```

```
(75589, 12)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 75589 entries, 0 to 75588
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   booking_id            75589 non-null  object
 1   property_id           75589 non-null  int64
 2   booking_date          75589 non-null  object
 3   check_in_date         75589 non-null  object
 4   checkout_date         75589 non-null  object
 5   no_guests             75589 non-null  int64
 6   room_category         75589 non-null  object
 7   booking_platform      75589 non-null  object
 8   ratings_given         31559 non-null  float64
 9   booking_status        75589 non-null  object
10   revenue_generated     75588 non-null  float64
11   revenue_realized      75588 non-null  float64
dtypes: float64(3), int64(2), object(7)
memory usage: 6.9+ MB
```

```
data = data.drop_duplicates()
```

```
data.shape
```

```
(75589, 12)
```

```
data.isnull().sum()
```

```
booking_id            0
property_id           0
booking_date          0
check_in_date         0
checkout_date         0
no_guests             0
room_category         0
booking_platform      0
ratings_given        44030
booking_status        0
revenue_generated     1
revenue_realized     1
dtype: int64
```

```
df = data.dropna()
```

```
df.isnull().sum()
```

```
booking_id      0
property_id     0
booking_date    0
check_in_date   0
checkout_date   0
no_guests       0
room_category   0
booking_platform 0
ratings_given   0
booking_status  0
revenue_generated 0
revenue_realized 0
dtype: int64
```

```
df.to_csv('fact_bookings.csv')
```

---

[Colab paid products](#) - [Cancel contracts here](#)

