

flight-booking-project

May 17, 2024

```
[ ]: import pandas as pd
```

```
[ ]: df = pd.read_csv("/content/drive/MyDrive/customer_booking123.csv")
```

```
[ ]: df.head()
```

```
[ ]:      num_passengers sales_channel trip_type purchase_lead length_of_stay \
0              2      Internet RoundTrip          262          19
1              1      Internet RoundTrip          112          20
2              2      Internet RoundTrip          243          22
3              1      Internet RoundTrip           96          31
4              2      Internet RoundTrip           68          22
```

```
      flight_hour flight_day route booking_origin wants_extra_baggage \
0              7      Sat AKLDEL New Zealand          1
1              3      Sat AKLDEL New Zealand          0
2             17      Wed AKLDEL India          1
3              4      Sat AKLDEL New Zealand          0
4             15      Wed AKLDEL India          1
```

```
      wants_preferred_seat wants_in_flight_meals flight_duration \
0              0              0          5.52
1              0              0          5.52
2              1              0          5.52
3              0              1          5.52
4              0              1          5.52
```

```
      booking_complete
0              0
1              0
2              0
3              0
4              0
```

```
[ ]: df.tail()
```

```
[ ]:      num_passengers sales_channel trip_type purchase_lead length_of_stay \
49995          2      Internet RoundTrip          27          6
49996          1      Internet RoundTrip          111          6
49997          1      Internet RoundTrip          24          6
49998          1      Internet RoundTrip          15          6
49999          1      Internet RoundTrip          19          6

      flight_hour flight_day route booking_origin wants_extra_baggage \
49995          9      Sat PERPNH      Australia          1
49996          4      Sun PERPNH      Australia          0
49997         22      Sat PERPNH      Australia          0
49998         11      Mon PERPNH      Australia          1
49999         10      Thu PERPNH      Australia          0

      wants_preferred_seat wants_in_flight_meals flight_duration \
49995          0          1          5.62
49996          0          0          5.62
49997          0          1          5.62
49998          0          1          5.62
49999          1          0          5.62

      booking_complete
49995          0
49996          0
49997          0
49998          0
49999          0
```

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   num_passengers         50000 non-null  int64
1   sales_channel          50000 non-null  object
2   trip_type              50000 non-null  object
3   purchase_lead          50000 non-null  int64
4   length_of_stay         50000 non-null  int64
5   flight_hour            50000 non-null  int64
6   flight_day             50000 non-null  object
7   route                  50000 non-null  object
8   booking_origin         50000 non-null  object
9   wants_extra_baggage    50000 non-null  int64
10  wants_preferred_seat   50000 non-null  int64
11  wants_in_flight_meals  50000 non-null  int64
```

```

12 flight_duration      50000 non-null float64
13 booking_complete     50000 non-null int64
dtypes: float64(1), int64(8), object(5)
memory usage: 5.3+ MB

```

```
[ ]: df.describe()
```

```
[ ]:
      num_passengers  purchase_lead  length_of_stay  flight_hour \
count      50000.000000      50000.000000      50000.000000      50000.000000
mean           1.591240          84.940480          23.04456          9.06634
std            1.020165          90.451378          33.88767          5.41266
min            1.000000           0.000000           0.00000          0.00000
25%            1.000000          21.000000           5.00000          5.00000
50%            1.000000          51.000000          17.00000          9.00000
75%            2.000000         115.000000          28.00000         13.00000
max            9.000000         867.000000         778.00000         23.00000

      wants_extra_baggage  wants_preferred_seat  wants_in_flight_meals \
count      50000.000000      50000.000000      50000.000000
mean           0.668780           0.296960           0.427140
std            0.470657           0.456923           0.494668
min            0.000000           0.000000           0.000000
25%            0.000000           0.000000           0.000000
50%            1.000000           0.000000           0.000000
75%            1.000000           1.000000           1.000000
max            1.000000           1.000000           1.000000

      flight_duration  booking_complete
count      50000.000000      50000.000000
mean           7.277561          0.149560
std            1.496863          0.356643
min            4.670000          0.000000
25%            5.620000          0.000000
50%            7.570000          0.000000
75%            8.830000          0.000000
max            9.500000          1.000000

```

```
[ ]: df.shape
```

```
[ ]: (50000, 14)
```

```
[ ]: df.columns
```

```
[ ]: Index(['num_passengers', 'sales_channel', 'trip_type', 'purchase_lead',
          'length_of_stay', 'flight_hour', 'flight_day', 'route',
          'booking_origin', 'wants_extra_baggage', 'wants_preferred_seat',
          'wants_in_flight_meals', 'flight_duration', 'booking_complete'],

```

```
dtype='object')
```

```
[ ]: df.dtypes
```

```
[ ]: num_passengers      int64
      sales_channel      object
      trip_type          object
      purchase_lead      int64
      length_of_stay     int64
      flight_hour        int64
      flight_day         object
      route             object
      booking_origin     object
      wants_extra_baggage int64
      wants_preferred_seat int64
      wants_in_flight_meals int64
      flight_duration    float64
      booking_complete   int64
      dtype: object
```

```
[ ]: df.isnull()
```

```
[ ]:      num_passengers  sales_channel  trip_type  purchase_lead  \
0          False          False          False          False
1          False          False          False          False
2          False          False          False          False
3          False          False          False          False
4          False          False          False          False
...          ...          ...          ...          ...
49995       False          False          False          False
49996       False          False          False          False
49997       False          False          False          False
49998       False          False          False          False
49999       False          False          False          False

      length_of_stay  flight_hour  flight_day  route  booking_origin  \
0          False          False          False  False          False
1          False          False          False  False          False
2          False          False          False  False          False
3          False          False          False  False          False
4          False          False          False  False          False
...          ...          ...          ...          ...
49995       False          False          False  False          False
49996       False          False          False  False          False
49997       False          False          False  False          False
49998       False          False          False  False          False
49999       False          False          False  False          False
```

	wants_extra_baggage	wants_preferred_seat	wants_in_flight_meals	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
...	
49995	False	False	False	
49996	False	False	False	
49997	False	False	False	
49998	False	False	False	
49999	False	False	False	

	flight_duration	booking_complete
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...
49995	False	False
49996	False	False
49997	False	False
49998	False	False
49999	False	False

[50000 rows x 14 columns]

```
[ ]: df.nunique()
```

```
[ ]: num_passengers          9
      sales_channel          2
      trip_type              3
      purchase_lead        470
      length_of_stay       335
      flight_hour          24
      flight_day           7
      route               799
      booking_origin       104
      wants_extra_baggage   2
      wants_preferred_seat  2
      wants_in_flight_meals 2
      flight_duration       21
      booking_complete      2
      dtype: int64
```

```
[ ]: df['num_passengers'].value_counts()
```

```
[ ]: num_passengers
1    31352
2    12804
3     2941
4     1784
5      561
6      289
7      114
8       89
9       66
Name: count, dtype: int64
```

```
[ ]: df.isnull().sum()
```

```
[ ]: num_passengers      0
sales_channel           0
trip_type               0
purchase_lead           0
length_of_stay          0
flight_hour             0
flight_day              0
route                   0
booking_origin          0
wants_extra_baggage     0
wants_preferred_seat    0
wants_in_flight_meals   0
flight_duration         0
booking_complete        0
dtype: int64
```

```
[ ]: df.duplicated().sum()
```

```
[ ]: 719
```

```
[ ]: df.drop_duplicates(inplace=True)
```

```
[ ]: df.duplicated().sum()
```

```
[ ]: 0
```

```
[ ]: df.dtypes
```

```
[ ]: num_passengers      int64
sales_channel          object
trip_type              object
```

purchase_lead	int64
length_of_stay	int64
flight_hour	int64
flight_day	object
route	object
booking_origin	object
wants_extra_baggage	int64
wants_preferred_seat	int64
wants_in_flight_meals	int64
flight_duration	float64
booking_complete	int64
dtype:	object