```
from google.colab import files
uploaded = files.upload()   # choose your Bookings-100000-Rows.xlsx file
```

```
Choose Files   No file chosen
```

```
import pandas as pd

# Replace the filename with the uploaded one if different
df = pd.read_excel("Bookings-100000-Rows.xlsx")

# Show first 5 rows
print(df.head())
```

```
                 Date      Time      Booking_ID       Booking_Status  \
0 2024-07-26 14:00:00  14:00:00  CNR7153255142     Canceled by Driver
1 2024-07-25 22:20:00  22:20:00  CNR2940424040                Success
2 2024-07-30 19:59:00  19:59:00  CNR2982357879                Success
3 2024-07-22 03:15:00  03:15:00  CNR2395710036   Canceled by Customer
4 2024-07-02 09:02:00  09:02:00  CNR1797421769                Success

  Customer_ID Vehicle_Type Pickup_Location Drop_Location  V_TAT  C_TAT  \
0   CID713523  Prime Sedan     Tumkur Road      RT Nagar    NaN    NaN
1   CID225428         Bike     Magadi Road       Varthur  203.0   30.0
2   CID270156    Prime SUV   Sahakar Nagar       Varthur  238.0  130.0
3   CID581320        eBike      HSR Layout   Vijayanagar    NaN    NaN
4   CID939555         Mini      Rajajinagar   Chamarajpet  252.0   80.0

                   Canceled_Rides_by_Customer        Canceled_Rides_by_Driver  \
0                                         NaN  Personal & Car related issue
1                                         NaN                             NaN
2                                         NaN                             NaN
3  Driver is not moving towards pickup location                             NaN
4                                         NaN                             NaN

  Incomplete_Rides Incomplete_Rides_Reason  Booking_Value Payment_Method  \
0              NaN                     NaN            444            NaN
1               No                     NaN            158           Cash
2               No                     NaN            386            UPI
3              NaN                     NaN            384            NaN
4               No                     NaN            822    Credit Card

   Ride_Distance  Driver_Ratings  Customer_Rating  \
0              0             NaN              NaN
1             13             4.1              4.0
2             40             4.2              4.8
3              0             NaN              NaN
4             45             4.0              3.0

                            Vehicle Images
0  https://cdn-icons-png.flaticon.com/128/14183/1...
1  https://cdn-icons-png.flaticon.com/128/9983/99...
2  https://cdn-icons-png.flaticon.com/128/9983/99...
3  https://cdn-icons-png.flaticon.com/128/6839/68...
4  https://cdn-icons-png.flaticon.com/128/3202/32...
```

```
print(df.columns)
```

```
Index(['Date', 'Time', 'Booking_ID', 'Booking_Status', 'Customer_ID',
       'Vehicle_Type', 'Pickup_Location', 'Drop_Location', 'V_TAT', 'C_TAT',
       'Canceled_Rides_by_Customer', 'Canceled_Rides_by_Driver',
       'Incomplete_Rides', 'Incomplete_Rides_Reason', 'Booking_Value',
       'Payment_Method', 'Ride_Distance', 'Driver_Ratings', 'Customer_Rating',
       'Vehicle Images'],
      dtype='object')
```

```
print(df.shape)
```

```
(103024, 20)
```

```
print(df.info)
```

```
<bound method DataFrame.info of                       Date      Time      Booking_ID       Booking_Status  \
0       2024-07-26 14:00:00  14:00:00  CNR7153255142     Canceled by Driver
1       2024-07-25 22:20:00  22:20:00  CNR2940424040                Success
2       2024-07-30 19:59:00  19:59:00  CNR2982357879                Success
3       2024-07-22 03:15:00  03:15:00  CNR2395710036   Canceled by Customer
4       2024-07-02 09:02:00  09:02:00  CNR1797421769                Success
...                    ...       ...            ...                    ...
103019  2024-07-31 09:06:00  09:06:00  CNR9488489435                Success
103020  2024-07-31 15:12:00  15:12:00  CNR3151743100                Success
```

```
103021  2024-07-31 13:59:00  13:59:00  CNR1286151233         Success
103022  2024-07-31 14:56:00  14:56:00  CNR2027162035         Success
103023  2024-07-31 13:57:00  13:57:00  CNR9770709721         Success

        Customer_ID Vehicle_Type    Pickup_Location  Drop_Location  V_TAT  \
0         CID713523  Prime Sedan        Tumkur Road       RT Nagar    NaN
1         CID225428         Bike        Magadi Road        Varthur  203.0
2         CID270156    Prime SUV      Sahakar Nagar        Varthur  238.0
3         CID581320        eBike         HSR Layout    Vijayanagar    NaN
4         CID939555         Mini        Rajajinagar     Chamarajpet  252.0
...             ...          ...                ...            ...    ...
103019    CID371654   Prime Plus      Richmond Town        Varthur  245.0
103020    CID334158         Auto        Vijayanagar  Richmond Town   84.0
103021    CID113188    Prime SUV  Bannerghatta Road       JP Nagar   35.0
103022    CID118301        eBike        Indiranagar    Magadi Road  210.0
103023    CID217959         Auto             Ulsoor         Hennur  175.0

         C_TAT                 Canceled_Rides_by_Customer  \
0          NaN                                        NaN
1         30.0                                        NaN
2        130.0                                        NaN
3          NaN  Driver is not moving towards pickup location
4         80.0                                        NaN
...        ...                                        ...
103019    35.0                                        NaN
103020   145.0                                        NaN
103021    75.0                                        NaN
103022   140.0                                        NaN
103023   125.0                                        NaN

           Canceled_Rides_by_Driver Incomplete_Rides Incomplete_Rides_Reason  \
0         Personal & Car related issue              NaN                     NaN
1                               NaN              No                     NaN
2                               NaN              No                     NaN
3                               NaN             NaN                     NaN
4                               NaN              No                     NaN
...                             ...             ...                     ...
103019                          NaN              No                     NaN
103020                          NaN              No                     NaN
103021                          NaN              No                     NaN
103022                          NaN              No                     NaN
103023                          NaN              No                     NaN

        Booking_Value Payment_Method  Ride_Distance  Driver_Ratings  \
0                 444            NaN              0             NaN
1                 158           Cash             13             4.1
2                 386            UPI             40             4.2
3                 384            NaN              0             NaN
4                 822    Credit Card             45             4.0
```

```
print(df.describe())
```

```
                         Date         V_TAT         C_TAT  \
count                  103024  63967.000000  63967.000000
mean   2024-07-16 11:31:38.879678720    170.876952     84.873372
min       2024-07-01 00:00:00     35.000000     25.000000
25%       2024-07-08 18:41:00     98.000000     55.000000
50%       2024-07-16 11:23:00    168.000000     85.000000
75%       2024-07-24 05:18:00    238.000000    115.000000
max       2024-07-31 23:58:00    308.000000    145.000000
std                       NaN     80.803640     36.005100

       Booking_Value  Ride_Distance  Driver_Ratings  Customer_Rating
count  103024.000000  103024.000000    63967.000000     63967.000000
mean      548.751883      14.189927        3.997457         3.998313
min       100.000000       0.000000        3.000000         3.000000
25%       242.000000       0.000000        3.500000         3.500000
50%       386.000000       8.000000        4.000000         4.000000
75%       621.000000      26.000000        4.500000         4.500000
max      2999.000000      49.000000        5.000000         5.000000
std       536.541221      15.776270        0.576834         0.578957
```

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

	0
Date	0
Time	0
Booking_ID	0
Booking_Status	0
Customer_ID	0
Vehicle_Type	0
Pickup_Location	0
Drop_Location	0
V_TAT	39057
C_TAT	39057
Canceled_Rides_by_Customer	92525
Canceled_Rides_by_Driver	84590
Incomplete_Rides	39057
Incomplete_Rides_Reason	99098
Booking_Value	0
Payment_Method	39057
Ride_Distance	0
Driver_Ratings	39057
Customer_Rating	39057
Vehicle Images	0

dtype: int64

```
df.columns=df.columns.str.lower().str.replace(" ","_")
print(df.columns)
```

```
Index(['date', 'time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'v_tat', 'c_tat',
       'canceled_rides_by_customer', 'canceled_rides_by_driver',
       'incomplete_rides', 'incomplete_rides_reason', 'booking_value',
       'payment_method', 'ride_distance', 'driver_ratings', 'customer_rating',
       'vehicle_images'],
      dtype='object')
```

```
print(df.duplicated().sum())
df=df.drop_duplicates()
print(df.shape)
```

```
0
(103024, 20)
```

```
df['pickup_time'] = pd.to_datetime(
    df['date'].astype(str) + ' ' + df['time'].astype(str),
    errors='coerce'
)
```

```
/tmp/ipython-input-3472887982.py:1: UserWarning: Could not infer format, so each element will be parsed individually, falling b
  df['pickup_time'] = pd.to_datetime(
```

```
df['pickup_time'].head()
```

	pickup_time
0	2024-07-26 14:00:00
1	2024-07-25 22:20:00
2	2024-07-30 19:59:00
3	2024-07-22 03:15:00
4	2024-07-02 09:02:00

dtype: datetime64[ns]

```python
# Check how many missing values are there in each column
print(df.isnull().sum())

# Example: Fill missing driver ratings with average rating
df['driver_ratings'] = df['driver_ratings'].fillna(df['driver_ratings'].mean())

# Example: Drop rows where booking_id or customer_id is missing (because they are important)
df = df.dropna(subset=['booking_id', 'customer_id'])

# Again check if missing values are handled
print(df.isnull().sum())
```

```
date                        0
time                        0
booking_id                  0
booking_status              0
customer_id                 0
vehicle_type                0
pickup_location             0
drop_location               0
v_tat                   39057
c_tat                   39057
canceled_rides_by_customer  92525
canceled_rides_by_driver    84590
incomplete_rides        39057
incomplete_rides_reason 99098
booking_value               0
payment_method          39057
ride_distance               0
driver_ratings          39057
customer_rating         39057
vehicle_images              0
pickup_time                 0
dtype: int64
date                        0
time                        0
booking_id                  0
booking_status              0
customer_id                 0
vehicle_type                0
pickup_location             0
drop_location               0
v_tat                   39057
c_tat                   39057
canceled_rides_by_customer  92525
canceled_rides_by_driver    84590
incomplete_rides        39057
incomplete_rides_reason 99098
booking_value               0
payment_method          39057
ride_distance               0
driver_ratings              0
customer_rating         39057
vehicle_images              0
pickup_time                 0
dtype: int64
```

```python
# Remove extra spaces in text columns
df['pickup_location'] = df['pickup_location'].str.strip()
df['drop_location']   = df['drop_location'].str.strip()

# Make booking_status consistent (example: 'Completed', 'Cancelled')
df['booking_status'] = df['booking_status'].str.title()
```

```python
# Convert ride_distance, booking_value to numeric
df['ride_distance'] = pd.to_numeric(df['ride_distance'], errors='coerce')
df['booking_value'] = pd.to_numeric(df['booking_value'], errors='coerce')
```

```python
# Convert ratings to numeric (just in case they are not)
df['driver_ratings']  = pd.to_numeric(df['driver_ratings'], errors='coerce')
df['customer_rating'] = pd.to_numeric(df['customer_rating'], errors='coerce')
```

```python
# Extract date features from pickup_time
df['year']    = df['pickup_time'].dt.year
df['month']   = df['pickup_time'].dt.month
df['day']     = df['pickup_time'].dt.day
df['weekday'] = df['pickup_time'].dt.day_name()
df['hour']    = df['pickup_time'].dt.hour

print(df[['pickup_time','year','month','day','weekday','hour']].head())
```

```
         pickup_time  year  month  day   weekday  hour
0 2024-07-26 14:00:00  2024      7   26    Friday    14
1 2024-07-25 22:20:00  2024      7   25  Thursday    22
2 2024-07-30 19:59:00  2024      7   30   Tuesday    19
3 2024-07-22 03:15:00  2024      7   22    Monday     3
4 2024-07-02 09:02:00  2024      7    2   Tuesday     9
```

```python
# Check duplicates based on booking_id
print("Duplicates:", df.duplicated(subset=['booking_id']).sum())

# Drop duplicates
df = df.drop_duplicates(subset=['booking_id'])
```

```
Duplicates: 0
```

```python
print(df.info())
print(df.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103024 entries, 0 to 103023
Data columns (total 26 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   date                      103024 non-null  datetime64[ns]
 1   time                      103024 non-null  object
 2   booking_id                103024 non-null  object
 3   booking_status            103024 non-null  object
 4   customer_id               103024 non-null  object
 5   vehicle_type              103024 non-null  object
 6   pickup_location           103024 non-null  object
 7   drop_location             103024 non-null  object
 8   v_tat                     63967 non-null   float64
 9   c_tat                     63967 non-null   float64
 10  canceled_rides_by_customer 10499 non-null  object
 11  canceled_rides_by_driver  18434 non-null   object
 12  incomplete_rides          63967 non-null   object
 13  incomplete_rides_reason   3926 non-null    object
 14  booking_value             103024 non-null  int64
 15  payment_method            63967 non-null   object
 16  ride_distance             103024 non-null  int64
 17  driver_ratings            103024 non-null  float64
 18  customer_rating           63967 non-null   float64
 19  vehicle_images            103024 non-null  object
 20  pickup_time               103024 non-null  datetime64[ns]
 21  year                      103024 non-null  int32
 22  month                     103024 non-null  int32
 23  day                       103024 non-null  int32
 24  weekday                   103024 non-null  object
 25  hour                      103024 non-null  int32
dtypes: datetime64[ns](2), float64(4), int32(4), int64(2), object(14)
memory usage: 18.9+ MB
None
                 date      time      booking_id        booking_status  \
0 2024-07-26 14:00:00  14:00:00  CNR7153255142    Canceled By Driver
1 2024-07-25 22:20:00  22:20:00  CNR2940424040               Success
2 2024-07-30 19:59:00  19:59:00  CNR2982357879               Success
3 2024-07-22 03:15:00  03:15:00  CNR2395710036  Canceled By Customer
4 2024-07-02 09:02:00  09:02:00  CNR1797421769               Success

  customer_id vehicle_type pickup_location drop_location  v_tat  c_tat  ... \
0   CID713523  Prime Sedan    Tumkur Road      RT Nagar    NaN    NaN  ...
1   CID225428         Bike    Magadi Road       Varthur  203.0   30.0  ...
2   CID270156    Prime SUV  Sahakar Nagar       Varthur  238.0  130.0  ...
3   CID581320        eBike     HSR Layout   Vijayanagar    NaN    NaN  ...
4   CID939555         Mini    Rajajinagar   Chamarajpet  252.0   80.0  ...

  ride_distance driver_ratings customer_rating  \
```

```
0          0   3.997457        NaN
1         13   4.100000        4.0
2         40   4.200000        4.8
3          0   3.997457        NaN
4         45   4.000000        3.0

                           vehicle_images        pickup_time  \
0   https://cdn-icons-png.flaticon.com/128/14183/1... 2024-07-26 14:00:00
1   https://cdn-icons-png.flaticon.com/128/9983/99    2024-07-25 22:20:00
```

```
0          0   3.997457        NaN
1         13   4.100000        4.0
2         40   4.200000        4.8
3          0   3.997457        NaN
4         45   4.000000        3.0

                           vehicle_images        pickup_time  \
0   https://cdn-icons-png.flaticon.com/128/14183/1... 2024-07-26 14:00:00
1   https://cdn-icons-png.flaticon.com/128/9983/99    2024-07-25 22:20:00
```