

Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1pBjbsZU5Q42_b2te9Tv#scrollTo=nli_NAnaNvnH

Google Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

- sample_data
- Data_Train.csv

+ Code + Text

```
[7] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

[11] data=pd.read_csv("Data_Train.csv")

[12] data.head()
```

	Airline	Date_of_Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24-03-2019	Banglore	New Delhi	BLR → DEL	22:20	22-03-2023 01:10	2h 50m	non-stop	No info	3897
1	Air India	01-05-2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	09-06-2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	10-06-2023 04:25	19h	2 stops	No info	13882
3	IndiGo	12-05-2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01-03-2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302

```
[13] data.shape
(10683, 11)

[14] data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   Airline             10683 non-null  object
 1   Date_of_Journey     10683 non-null  object
 2   Source              10683 non-null  object
 3   Destination         10683 non-null  object
 4   Route               10683 non-null  object
 5   Dep_Time            10683 non-null  object
 6   Arrival_Time        10683 non-null  object
 7   Duration            10683 non-null  object
 8   Total_Stops         10683 non-null  object
 9   Additional_Info     10683 non-null  object
10   Price              10683 non-null  int64
dtypes: int64(1), object(10)
memory usage: 918.2+ KB
```

0s completed at 7:22 AM

28°C Partly sunny 07:25 AM 19-04-2023

Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1pBjbsZU5Q42_b2te9Tv#scrollTo=nli_NAnaNvnH

Google Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

- sample_data
- Data_Train.csv

+ Code + Text

```
(10683, 11)

[14] data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   Airline             10683 non-null  object
 1   Date_of_Journey     10683 non-null  object
 2   Source              10683 non-null  object
 3   Destination         10683 non-null  object
 4   Route               10683 non-null  object
 5   Dep_Time            10683 non-null  object
 6   Arrival_Time        10683 non-null  object
 7   Duration            10683 non-null  object
 8   Total_Stops         10683 non-null  object
 9   Additional_Info     10683 non-null  object
10   Price              10683 non-null  int64
dtypes: int64(1), object(10)
memory usage: 918.2+ KB
```

```
[15] data.isnull().sum()
```

```
Airline      0
Date_of_Journey  0
Source        0
Destination   0
Route         1
Dep_Time      0
Arrival_Time  0
Duration      0
Total_Stops   1
Additional_Info 0
Price         0
dtype: int64
```

0s completed at 7:22 AM

28°C Partly sunny 07:25 AM 19-04-2023

Colaboratory interface showing a Jupyter Notebook with the following code and output:

```
[15] data.isnull().sum()
Airline      0
Date_of_Journey  0
Source        0
Destination   0
Route         1
Dep_Time      0
Arrival_Time  0
Duration      0
Total_Stops   1
Additional_Info 0
Price         0
dtype: int64
```

```
[16] data.dropna(inplace=True)
```

```
[17] data.isnull().sum()
Airline      0
Date_of_Journey  0
Source        0
Destination   0
Route         0
Dep_Time      0
Arrival_Time  0
Duration      0
Total_Stops   0
Additional_Info 0
Price         0
dtype: int64
```

```
[18] category=['Airline','Source','Destination','Additional_Info']
category
['Airline', 'Source', 'Destination', 'Additional_Info']
```

The interface also shows a file explorer on the left with files like 'sample_data' and 'Data_Train.csv'. The bottom status bar indicates the notebook is completed at 7:22 AM.

Colaboratory interface showing a Jupyter Notebook with the following code and output:

```
[18] category=['Airline','Source','Destination','Additional_Info']
category
['Airline', 'Source', 'Destination', 'Additional_Info']
```

```
[19] for i in category:
    print(i, data[i].unique())

Airline ['Indigo' 'Air India' 'Jet Airways' 'SpiceJet' 'Multiple carriers' 'GoAir'
'Vistara' 'Air Asia' 'Vistara Premium economy' 'Jet Airways Business'
'Multiple carriers Premium economy' 'Trujet']
Source ['Bangalore' 'Kolkata' 'Delhi' 'Chennai' 'Mumbai']
Destination ['New Delhi' 'Bangalore' 'Cochin' 'Kolkata' 'Delhi' 'Hyderabad']
Additional_Info ['No info' 'In-flight meal not included' 'No check-in baggage included'
'1 Short layover' 'No Info' '1 Long layover' 'Change airports'
'Business class' 'Red-eye flight' '2 Long layover']
```

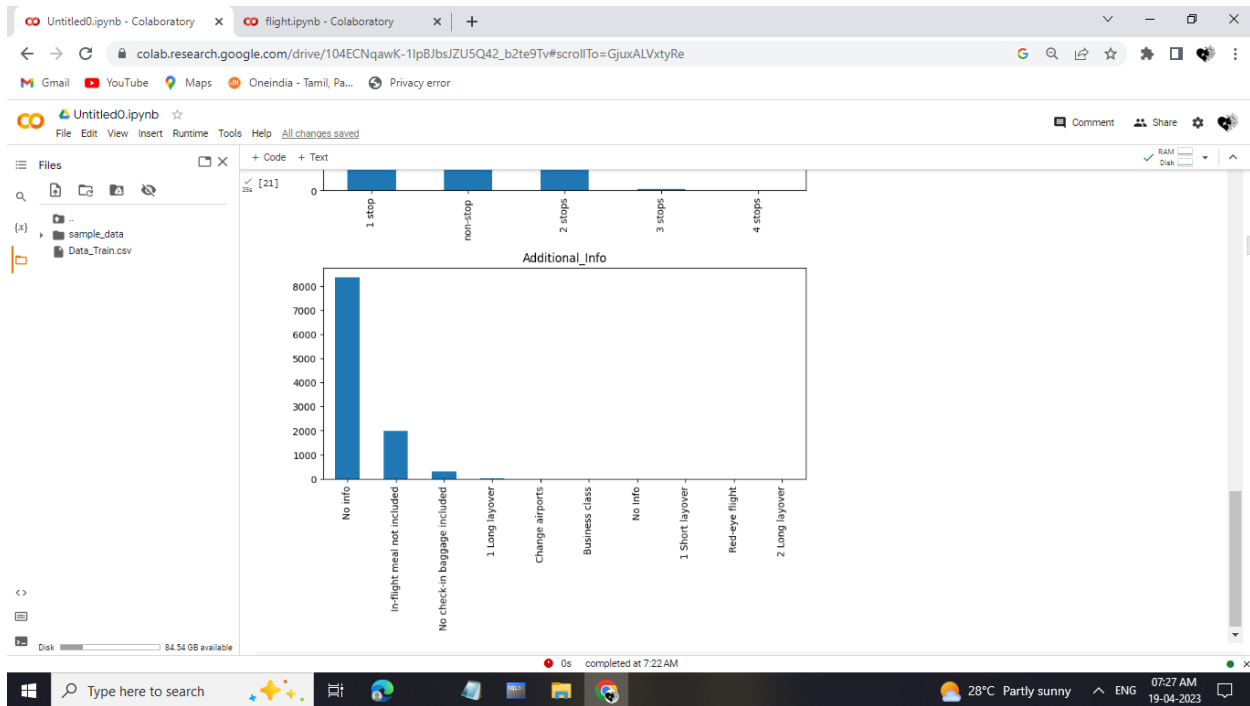
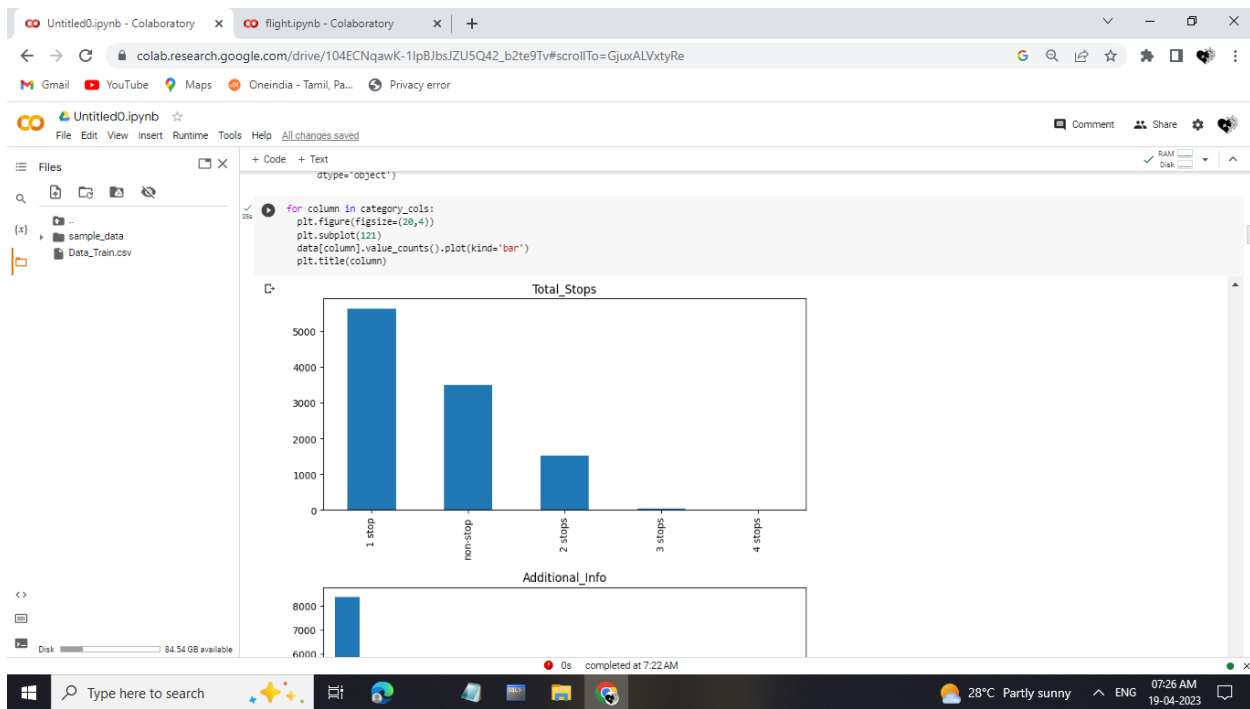
```
[20] category_cols=data.select_dtypes(include=['object']).columns
category_cols
Index(['Airline', 'Date_of_Journey', 'Source', 'Destination', 'Route',
'Dep_Time', 'Arrival_Time', 'Duration', 'Total_Stops',
'Additional_Info'],
      dtype='object')
```

```
[21] for column in category_cols:
    plt.figure(figsize=(20,4))
    plt.subplot(121)
    data[column].value_counts().plot(kind='bar')
    plt.title(column)
```

The output shows a bar chart titled 'Total_Stops' with the following data series:

Total_Stops	Count
0	10
1	10

The interface also shows a file explorer on the left with files like 'sample_data' and 'Data_Train.csv'. The bottom status bar indicates the notebook is completed at 7:22 AM.



Colaboratory interface showing a Jupyter Notebook with the following code:

```
data.Route=data.Route.str.split('->')
data.Route

0      [BLR -> DEL]
1      [CCU -> IXR -> BBI -> BLR]
2      [DEL -> LKO -> BOM -> COK]
3      [CCU -> NAG -> BLR]
4      [BLR -> NAG -> DEL]
...
10678   [CCU -> BLR]
10679   [CCU -> BLR]
10680   [BLR -> DEL]
10681   [BLR -> DEL]
10682   [DEL -> GOI -> BOM -> COK]
Name: Route, Length: 10682, dtype: object

[23] data['City1']=data.Route.str[0]
data['City2']=data.Route.str[1]
data['City3']=data.Route.str[2]
data['City4']=data.Route.str[3]
data['City5']=data.Route.str[4]
data['City6']=data.Route.str[5]

[24] data.Date_of_Journey=data.Date_of_Journey.str.split('/')
data.Date_of_Journey

0      [24-03-2019]
1      [01-06-2019]
2      [09-06-2019]
3      [12-05-2019]
4      [01-03-2019]
...
10678   [09-04-2019]
10679   [27-04-2019]
10680   [27-04-2019]
```

The interface includes a file explorer on the left showing 'sample_data' and 'Data_Train.csv'. The bottom status bar indicates '0s completed at 7:22 AM'.

Colaboratory interface showing a Jupyter Notebook with the following code:

```
[24] data.Date_of_Journey=data.Date_of_Journey.str.split('/')
data.Date_of_Journey

0      [24-03-2019]
1      [01-06-2019]
2      [09-06-2019]
3      [12-05-2019]
4      [01-03-2019]
...
10678   [09-04-2019]
10679   [27-04-2019]
10680   [27-04-2019]
10681   [01-03-2019]
10682   [09-04-2019]
Name: Date_of_Journey, Length: 10682, dtype: object

[25] data['Date']=data.Date_of_Journey.str[0]
data['Months']=data.Date_of_Journey.str[1]
data['Year']=data.Date_of_Journey.str[2]

[26] data.Dep_Time=data.Dep_Time.str.split(':')
data['Dep_Time_Hour']=data.Dep_Time.str[0]
data['Dep_Time_Mins']=data.Dep_Time.str[1]

[28] data.Arrival_Time=data.Arrival_Time.str.split(':')

[29] data['Arrival_data']=data.Arrival_Time.str[1]
data['Time_of_Arrival']=data.Arrival_Time.str[1]

[30] data.Duration=data.Duration.str.split('')
```

The interface includes a file explorer on the left showing 'sample_data' and 'Data_Train.csv'. The bottom status bar indicates '0s completed at 7:22 AM'.

Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1lpBjbsJZU5Q42_b2te9Tv#scrollTo=GjuxALVxyRe

Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

sample_data
Data_Train.csv

```
[26] data.Dep_Time=data.Dep_Time.str.split(':')

[27] data['Dep_Time_Hour']=data.Dep_Time.str[0]
data['Dep_Time_Mins']=data.Dep_Time.str[1]

[28] data.Arrival_Time=data.Arrival_Time.str.split(':')

[29] data['Arrival_data']=data.Arrival_Time.str[1]
data['Time_of_Arrival']=data.Arrival_Time.str[1]

[30] data.Duration=data.Duration.str.split(' ')

[31] data['Travel_Hours']=data.Duration.str[0]
data['Travel_Hours']=data['Travel_Hours'].str.split('h')
data['Travel_Hours']=data['Travel_Hours'].str[0]
data.Travel_Hours=data.Travel_Hours
data['Travel_Mins']=data.Duration.str[1]
data.Travel_Mins=data.Travel_Mins.str.split('m')
data.Travel_Mins=data.Travel_Mins.str[0]

[32] data.Total_Stops.replace('non_stop',0,inplace=True)
data.Total_Stops=data.Total_Stops.str.split(' ')
data.Total_Stops=data.Total_Stops.str[0]

[33] data.Additional_Info.unique()

array(['No info', 'In-flight meal not included',
       'No check-in baggage included', '1 short layover', 'No info',
       '1 Long layover', 'Change airports', 'Business class',
       'Red-eye flight', '2 Long layover'], dtype=object)
```

0s completed at 7:22 AM

Type here to search 28°C Partly sunny ENG 07:28 AM 19-04-2023

Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1lpBjbsJZU5Q42_b2te9Tv#scrollTo=UssjuhR5k7C

Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

sample_data
Data_Train.csv

```
[33] data.Additional_Info.unique()

array(['No info', 'In-flight meal not included',
       'No check-in baggage included', '1 short layover', 'No info',
       '1 Long layover', 'Change airports', 'Business class',
       'Red-eye flight', '2 Long layover'], dtype=object)

[34] data.Additional_Info.replace('No Info','No info',inplace=True)

data.isnull().sum()

Airline      0
Date_of_Journey  0
Source       0
Destination  0
Route        0
Dep_Time     0
Arrival_Time 0
Duration     0
Total_Stops  0
Additional_Info 0
Price        0
City1        0
City2      10682
City3      10682
City4      10682
City5      10682
City6      10682
Date         0
Months      10682
Year        10682
Dep_Time_Hour  0
Dep_Time_Mins 0
Arrival_data  0
Time_of_Arrival 0
Travel_Hours  0
Travel_Mins   0
dtype: int64
```

0s completed at 7:22 AM

Type here to search 28°C Partly sunny ENG 07:29 AM 19-04-2023

Google Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

RAM 100% Disk 100%

Files

sample_data Data_Train.csv

```
[35] data.drop(['Time_of_Arrival', 'Travel_Hours', 'Travel_Mins'], axis=1, inplace=True)

[36] data.drop(['City4', 'City5', 'City6'], axis=1, inplace=True)

[37] data.drop(['Date_of_Journey', 'Route', 'Dep_Time', 'Arrival_Time', 'Duration'], axis=1, inplace=True)
data.drop(['Time_of_Arrival'], axis=1, inplace=True)

[38] data.isnull().sum()

Airline      0
Source       0
Destination  0
Total_Stops  0
Additional_Info  0
Price        0
City1        0
City2      10682
City3      10682
Date         0
Months       0
Year         0
Dep_Time_Hour  0
Dep_Time_Mins  0
Arrival_data  0
Travel_Hours  0
Travel_Mins  0
dtype: int64

[40] data['City3'].fillna('None', inplace=True)
```

Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1pBJbsJZU5Q42_b2te9Tv#scrollTo=USsjuhR5k7C

Google Gmail YouTube Maps Oneindia - Tamil Pa... Privacy error

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

RAM 100% Disk 100%

Files

sample_data Data_Train.csv

```
[49] data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 10682 entries, 0 to 10682
Data columns (total 17 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Airline         10682 non-null  object
 1   Source          10682 non-null  object
 2   Destination     10682 non-null  object
 3   Total_Stops     10682 non-null  object
 4   Additional_Info  10682 non-null  object
 5   Price           10682 non-null  int64
 6   City1           10682 non-null  object
 7   City2           0 non-null      float64
 8   City3           10682 non-null  object
 9   Date            10682 non-null  object
10  Months          0 non-null      float64
11  Year            0 non-null      float64
12  Dep_Time_Hour   10682 non-null  object
13  Dep_Time_Mins   10682 non-null  object
14  Arrival_data    10682 non-null  object
15  Travel_Hours    10682 non-null  object
16  Travel_Mins     10682 non-null  object
dtypes: float64(3), int64(1), object(13)
memory usage: 1.5+ MB

[51] data[data['Travel_Hours'] == '5m']

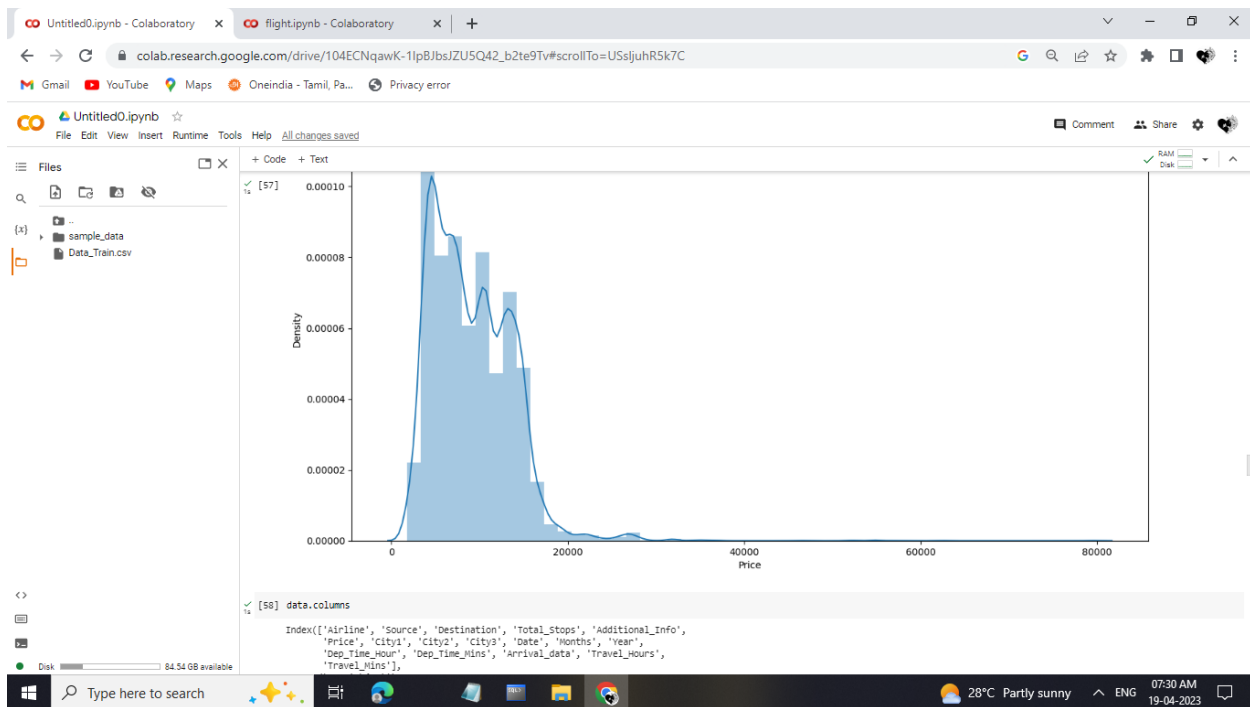
Airline Source Destination Total_Stops Additional_Info Price City1 City2 City3 Date Months Year Dep_Time_Hour Dep_Time_Mins Arrival_data Travel_Hours Travel_Mins

[52] data.drop(index=6474, inplace=True, axis=0)

[73] from sklearn.model_selection import RandomizedSearchCV
```

0s completed at 7:22 AM

28°C Partly sunny ENG 07:30 AM 19-04-2023



Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1lpBjbsJZU5Q42_b2te9Tv#scrollTo=USsjuhR5k7C

Gmail YouTube Maps Oneindia - Tamil, Pa... Privacy error

Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

Files

(x) sample_data Data_Train.csv

+ Code + Text

[58] Index(['Airline', 'Source', 'Destination', 'total_stops', 'Additional_Info',

```
       'Price', 'City1', 'City2', 'City3', 'Date', 'Months', 'Year',
       'Dep_Time_Hour', 'Dep_Time_Mins', 'Arrival_data', 'Travel_Hours',
       'Travel_Mins'],
      dtype='object')
```

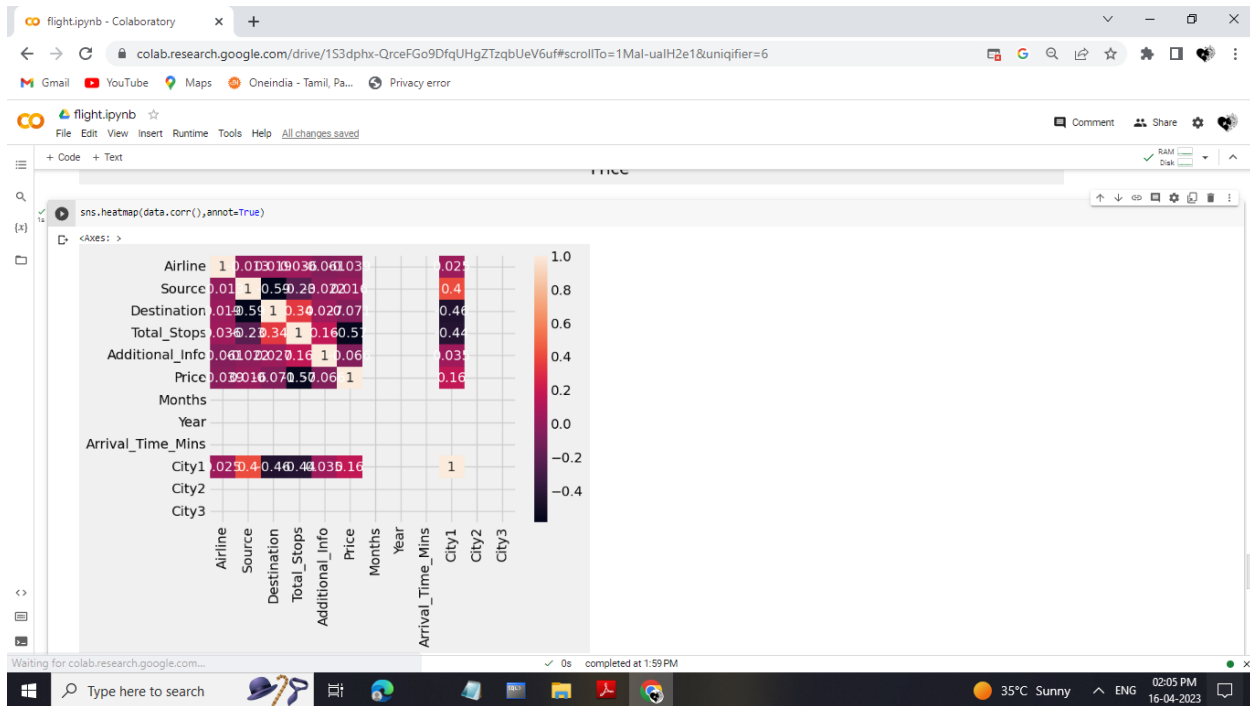

Colaboratory interface showing a Jupyter Notebook with the following code and output:

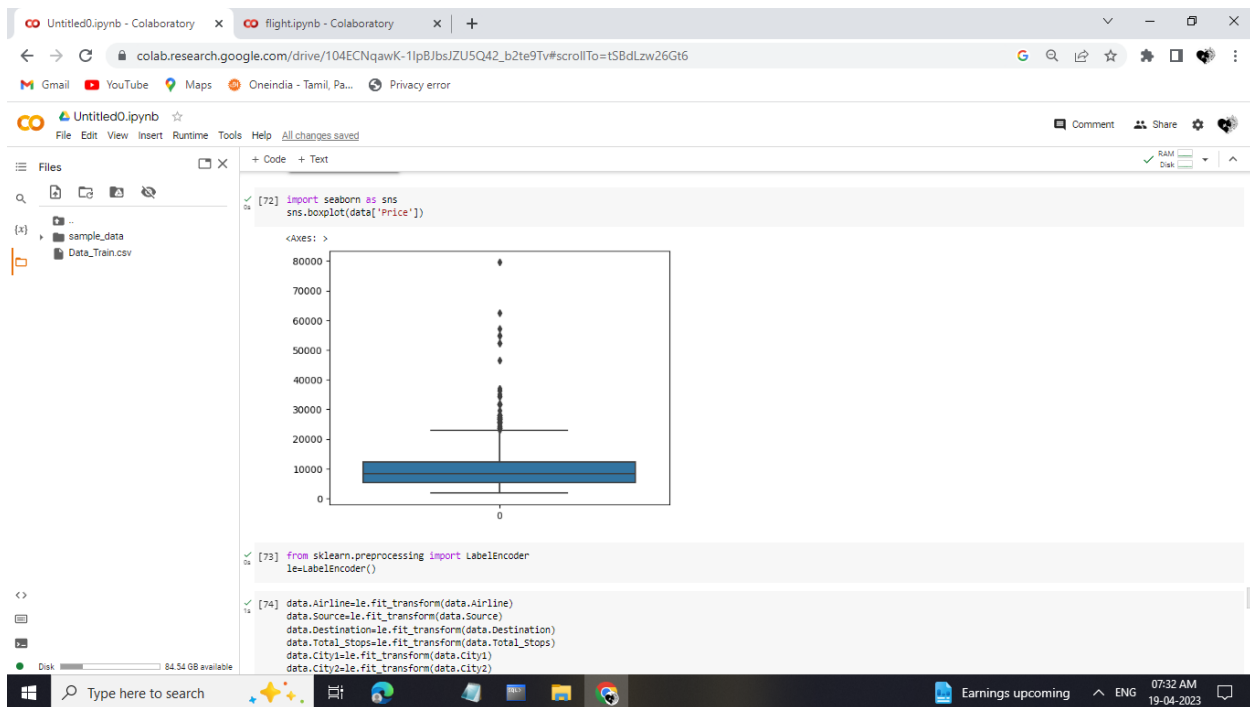
```
[62]: data[data.Price>5000]
data.head()
pd.set_option('display.max_columns',25)
data.head()
```

	Airline	Source	Destination	Total_Stops	Additional_Info	Price	City1	City2	City3	Date	Months	Year	Dep_Time_Hour	Dep_Time_Mins	Arrival_data	Travel_Hours	Travel_Mins
0	IndiGo	Banglore	New Delhi		No info	3897	BLR → DEL	NaN	None	24-03-2019	NaN	NaN	22	20	2		2
1	Air India	Kolkata	Banglore		No info	7662	CCU → IXR → BBI → BLR	NaN	None	01-05-2019	NaN	NaN	05	50	1		7
2	Jet Airways	Delhi	Cochin		No info	13882	DEL → LKO → BOM → COK	NaN	None	09-06-2019	NaN	NaN	09	25	1		1
3	IndiGo	Kolkata	Banglore		No info	6218	CCU → NAG → BLR	NaN	None	12-05-2019	NaN	NaN	18	05	2		5
4	IndiGo	Banglore	New Delhi		No info	13302	BLR → NAG → DEL	NaN	None	01-03-2019	NaN	NaN	16	50	2		4

```
[66]: data['Year'].max()
nan
```

System tray: 28°C Partly sunny, 07:31 AM 19-04-2023





Untitled0.ipynb - Colaboratory x flight.ipynb - Colaboratory x +

colab.research.google.com/drive/104ECNqawK-1lpBjbsJZU5Q42_b2te9Tv#scrollTo=P42YhmVOMHiM

Google Gmail YouTube Maps Oneindia - Tamil, Pa... Privacy error

Untitled0.ipynb File Edit View Insert Runtime Tools Help All changes saved Comment Share Settings

Files (x) sample_data Data_Train.csv

```
[73] from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
[74] data.Airline=le.fit_transform(data.Airline)
data.Source=le.fit_transform(data.Source)
data.Destination=le.fit_transform(data.Destination)
data.Total_Stops=le.fit_transform(data.Total_Stops)
data.City1=le.fit_transform(data.City1)
data.City2=le.fit_transform(data.City2)
data.City3=le.fit_transform(data.City3)
data.Additional_Info=le.fit_transform(data.Additional_Info)
data.head()
```

	Airline	Source	Destination	Total_Stops	Additional_Info	Price	City1	City2	City3	Date	Months	Year	Dep_Time_Hour	Dep_Time_Mins	Arrival_data	Travel_Hours	Travel_Mins
0	3	0	5	0	7	3897	18	0	0	24-03-2019	NaN	NaN	22	20	2	2	
1	1	3	0	0	7	7662	84	0	0	01-05-2019	NaN	NaN	05	50	1	7	
2	4	2	1	0	7	13882	118	0	0	09-06-2019	NaN	NaN	09	25	1	1	
3	3	3	0	0	7	6218	91	0	0	12-05-2019	NaN	NaN	18	05	2	5	
4	3	0	5	0	7	13302	29	0	0	01-03-2019	NaN	NaN	16	50	2	4	

```
[75] from sklearn.preprocessing import StandardScaler
ss=StandardScaler()
```

Type here to search Earnings upcoming ENG 07:33 AM 19-04-2023