OPTIMIZING FLIGHT BOOKING DECISIONS THROUGH MACHINE LEARNING

PRICE PREDICTIONS

INTRODUCTION

OVERVIEW

- People who work frequently travel through flight will have better knowledge on best discount and right time to buy the ticket.

- For the business purpose many airline companies change prices according to the seasons or time.

- They will increase the price when people travel more. Estimating the highest prices of the airline data for the route is collected with features such as Duration, Destination, Arrival and Departure.

- Features are taken from chosen dataset and in the price where in the airline price ticket costs vary overtime.

- We have implemented flight price prediction for users by using KNN, decision tree and random forest algorithms.

- Random Forest show the best accuracy of 80% for predicting the flight price. Also, we have done correlation tests and metrics for the statistical analysis.

PURPOSE

- ✓ Airline tickets are important document that confirm a passenger has a sent on a flight.

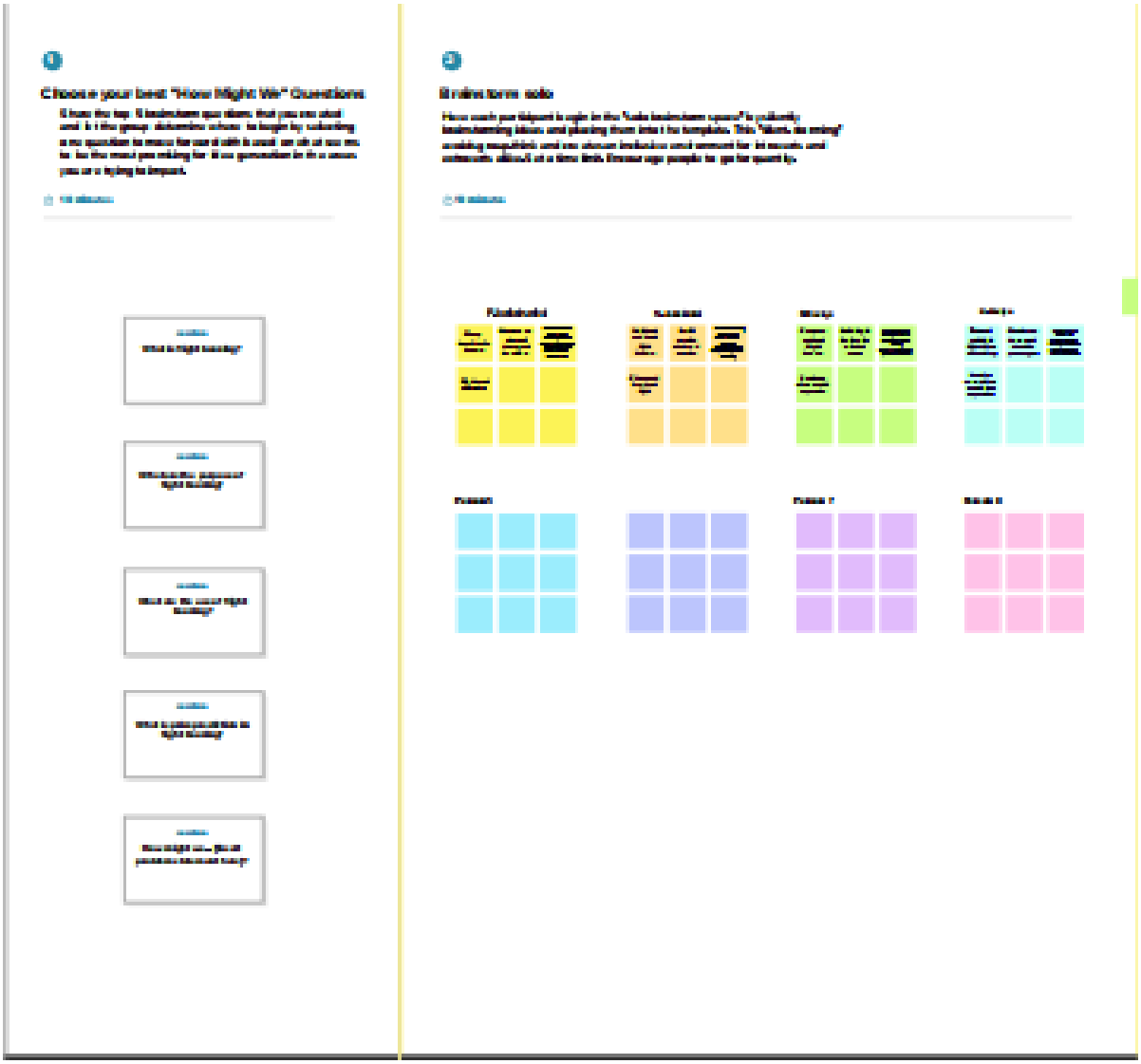
- ✓ The ticket includes important information about the passenger and the flight that they will take.

- ✓ The ticket is exchanged for a boarding pass during the check-in process, and this gives the passengers permission to board the plane.

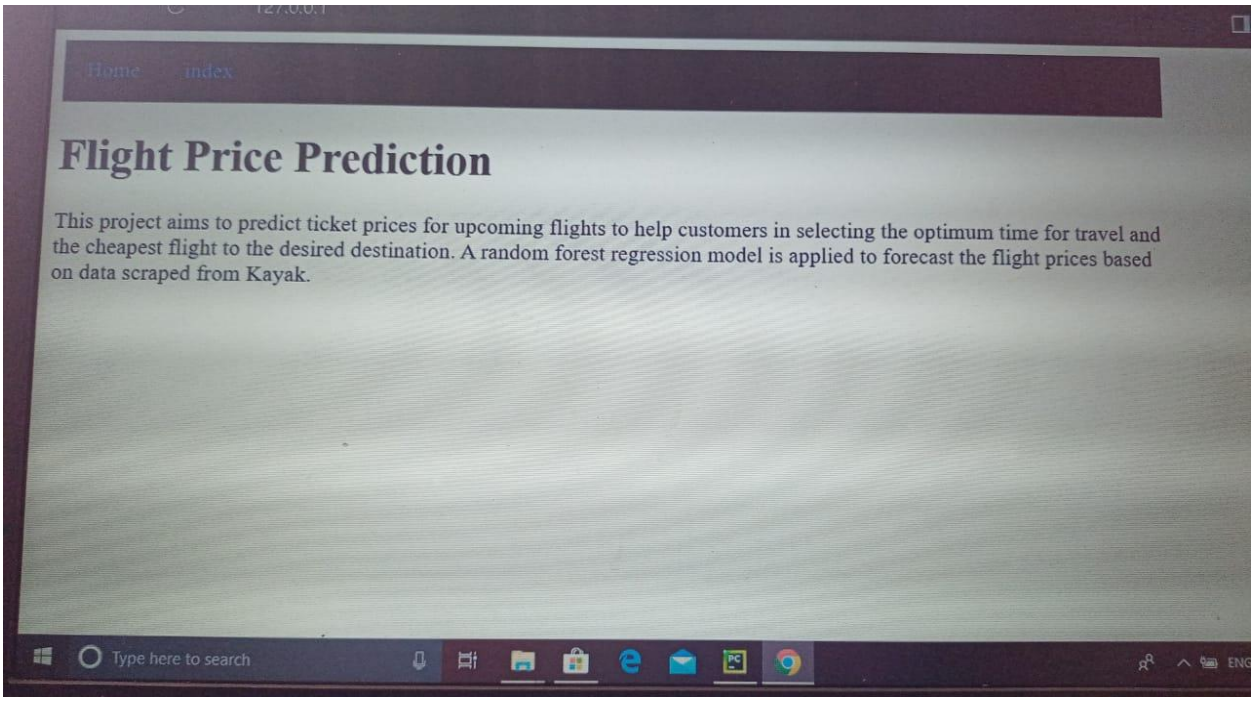
PROBLEM DEFINITION & DESIGN THINKING

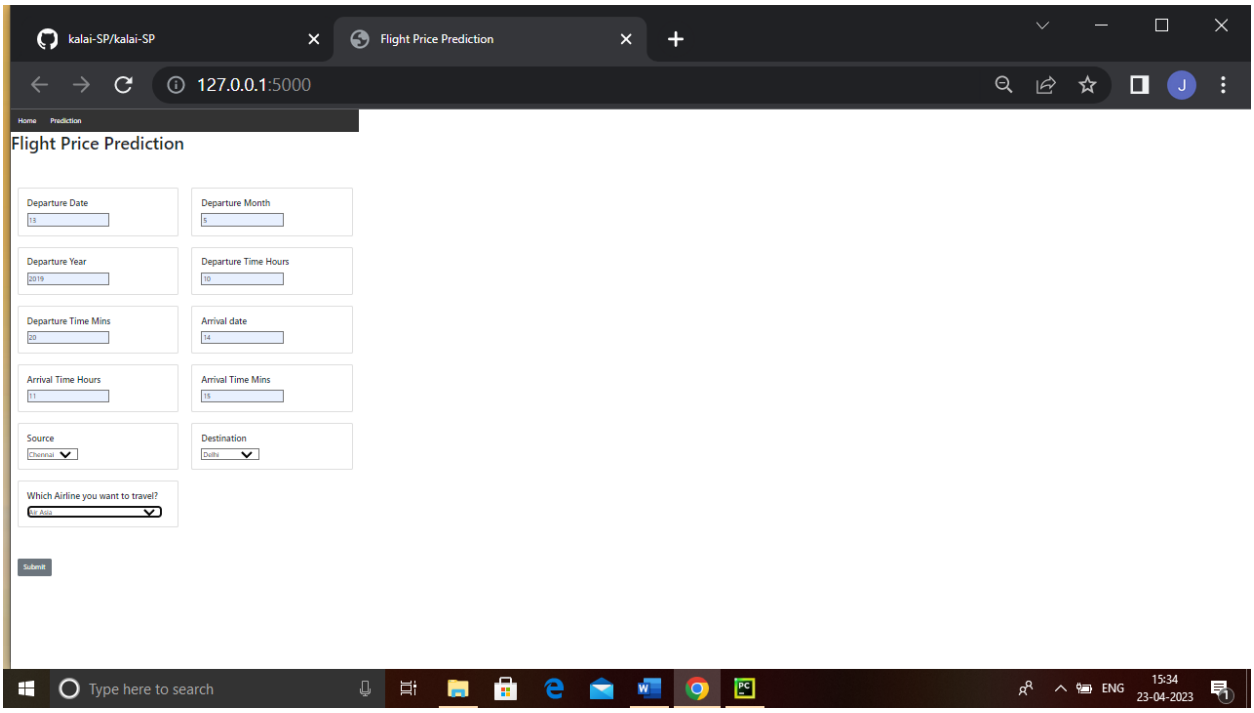
EMPATHY MAP

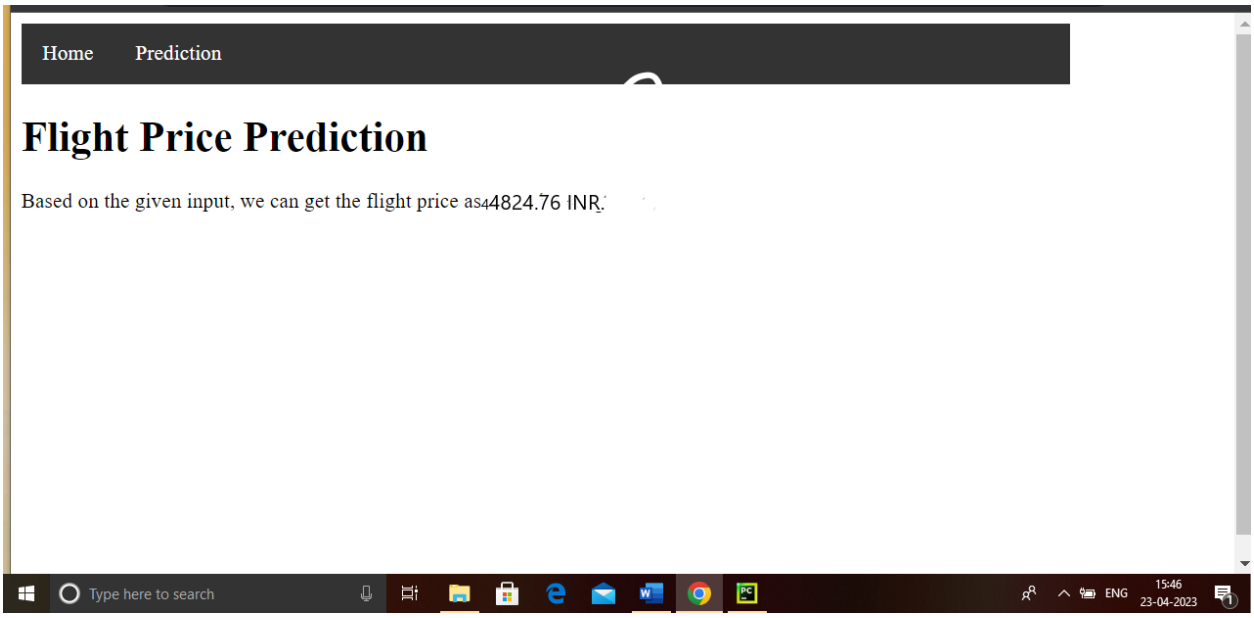
IDEATION & BRAINSTORMING MAP



RESULT


Home
index
Flight Price Prediction
This project aims to predict ticket prices for upcoming flights to help customers in selecting the optimum time for travel and the cheapest flight to the desired destination. A random forest regression model is applied to forecast the flight prices based on data scraped from Kayak.
Type here to search


kalai-SP/kalai-SP
Flight Price Prediction
127.0.0.1:5000
Home
Prediction
Flight Price Prediction
Departure Date
15
Departure Month
5
Departure Year
2019
Departure Time Hours
10
Departure Time Mins
20
Arrival date
14
Arrival Time Hours
11
Arrival Time Mins
15
Source
Chennai
Destination
Delhi
Which Airline you want to travel?
Air Asia
Submit
Type here to search
ENG
15:34
23-04-2023



ADVANTAGES

- They’re called budget airline for a reason lower rates that make for appealing flight fares for budget travelers.
- First remember these airline serve fewer routes.
- Speaking of routers. The best budget airline fly both national and international routes.
- To take advantages of the promotion make it a habit to always check a booking website for new promo fares and discount packages.

DISADVANTAGE

- Many budget airline don’t use major airports for takeoff and landing.
- Beware of hidden charges aide from the original fare advertised on the website.
- Another possible issue you might face with promotional tickets is that dates or destinations may be limited.
- Also the flight status might be delayed as well.

APPLICATION

- Fare Arena
- Kayak
- Air India
- IndiGo-Book flight ticket
- Kicoi.com-Book cheap & flights
- Vistara-India’s Best Airline
- Last minute flight Booking
- Expedia.

CONCLSUION

- ❖ Flight ticket booking app development latest trend in the market as the travel is rising globally, and people are replying app for easy trips. Machine learning algorithms are applied on the dataset to predict the dynamic fare of flight.

- ❖ This gives the predicted values of flight fare to get a flight ticket at minimum cost.

- ❖ Data is collected from the websites which sell the flight tickets so only limited information can be accessed, The values of R-squared obtained from the algorithm give the accuracy of the model.

- ❖ In the future. If more data could be accessed such as the current availability of seats. The predicted results will be more accurate. Finally. We have created the entire process of predicting an airline ticket and given a proof of our prediction based on the previous trends with our prediction.

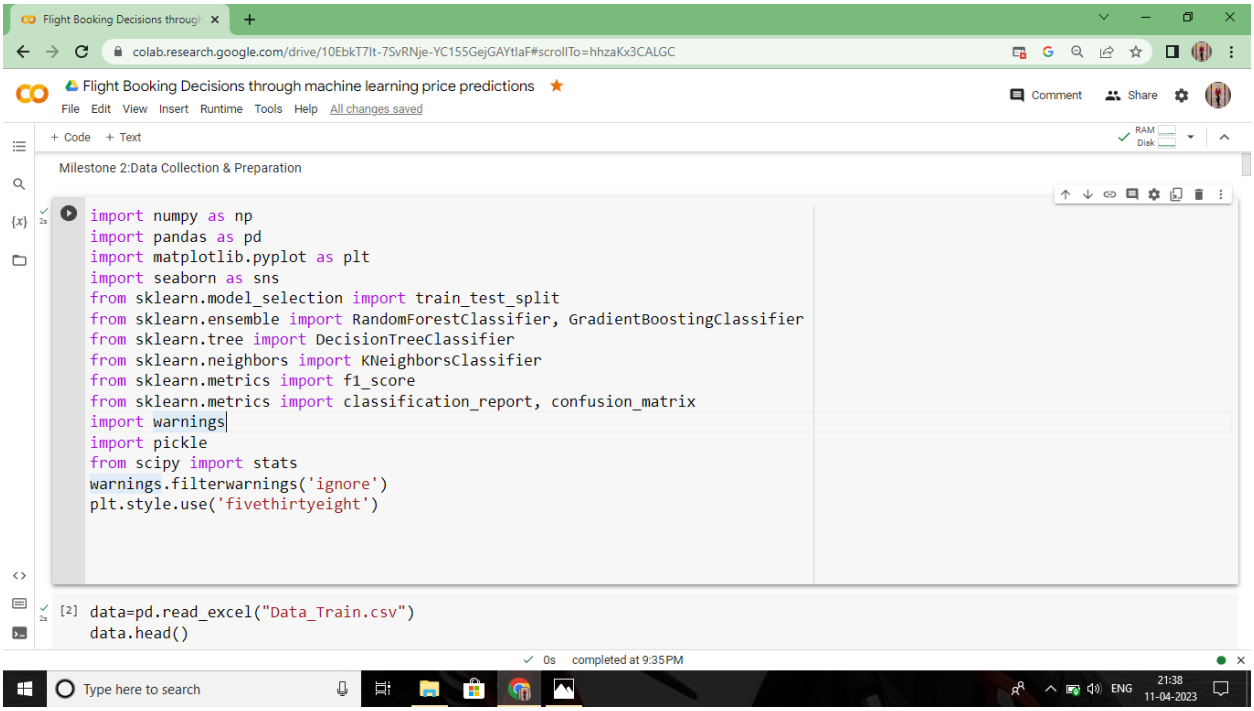
FUTURE SCOPE

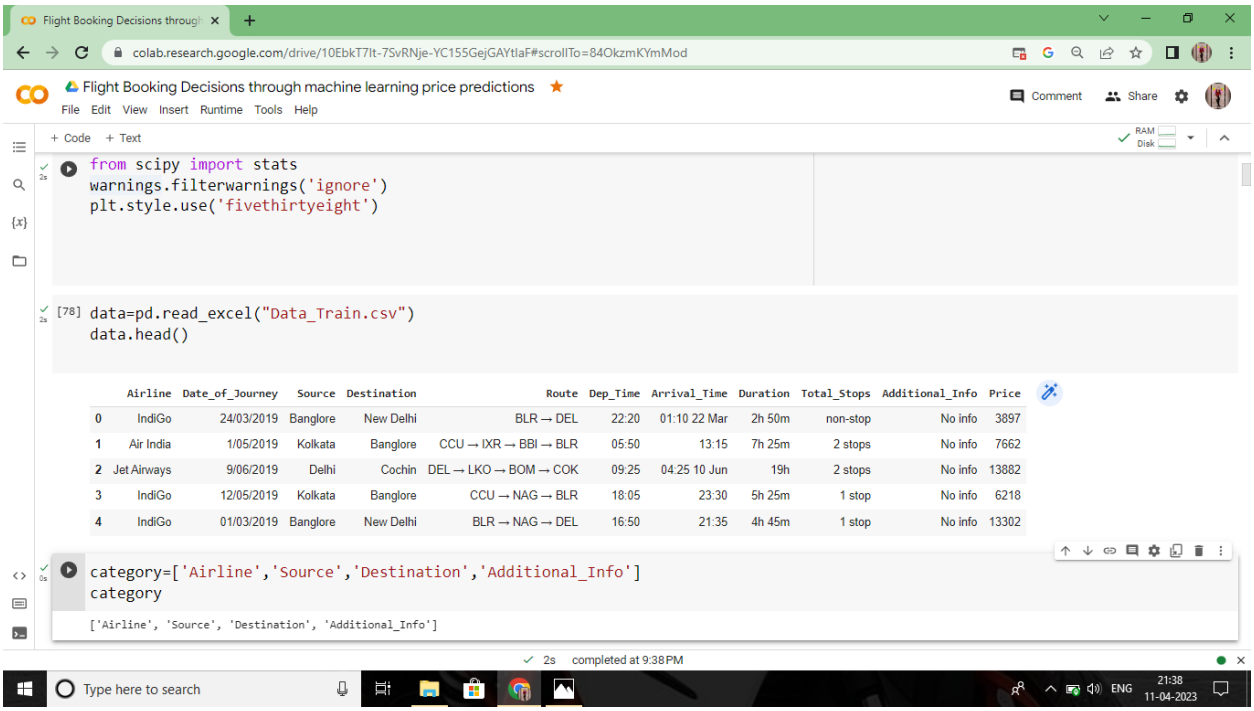
- Air ticketing is very dynamic and booming work sphere with great future scope on a global level.

- The Air ticketing industry is one of the industry is one of the largest and most profitable industries in the word and contributes substantially to the foreign exchange earned.

APPENDIX

A. Source Code





Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=YbQluVA2fE5e

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302

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

```
['Airline', 'Source', 'Destination', 'Additional_Info']
```

```
[4] 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 ['Banglore' 'Kolkata' 'Delhi' 'Chennai' 'Mumbai']
Destination ['New Delhi' 'Banglore' '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']
```

```
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')
```

2s completed at 9:38PM

Type here to search

ENG 21:39 11-04-2023

Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=YbQluVA2fE5e

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

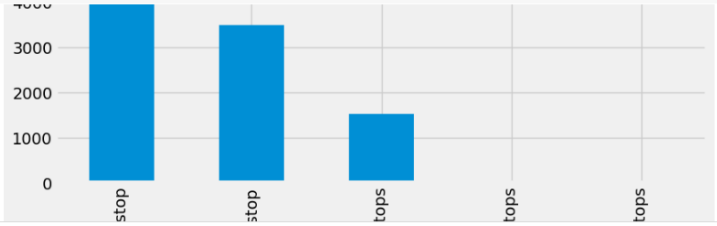
+ Code + Text

RAM Disk

```
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')
```

```
[6] #plotting a barchart for each of the categorical value
    for columns in category_cols:
      plt.figure(figsize=(20,4))
      plt.subplot(121)
      data[columns].value_counts().plot(kind='bar')
      plt.title(columns)
```



2s completed at 9:38PM

Type here to search

ENG 21:39 11-04-2023

Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=6ELk_1jfZ--

Flight Booking Decisions through machine learning price predictions

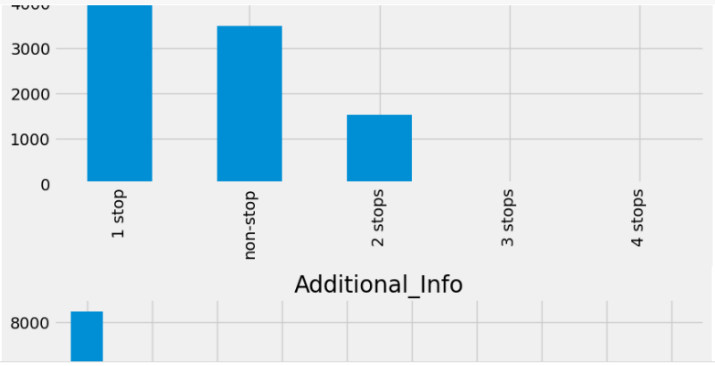
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

```
#plotting a barchart for each of the categorical value
for columns in category_cols:
  plt.figure(figsize=(20,4))
  plt.subplot(121)
  data[columns].value_counts().plot(kind='bar')
  plt.title(columns)
```



2s completed at 9:38 PM

Type here to search

ENG 21:39 11-04-2023

Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=6ELk_1jfZ--

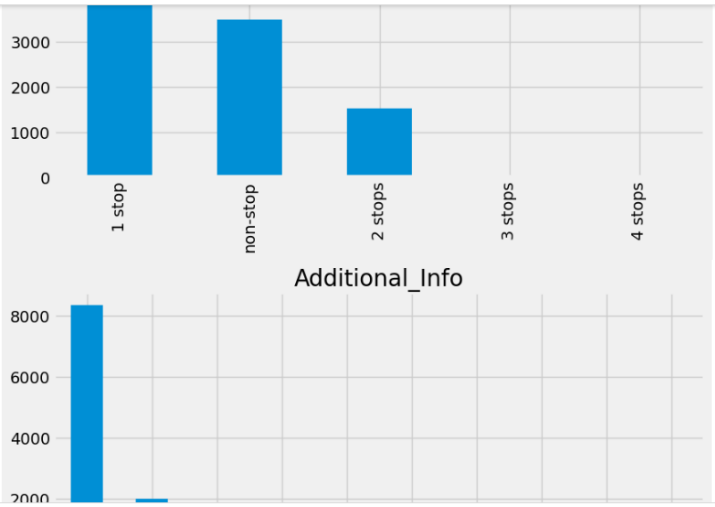
Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

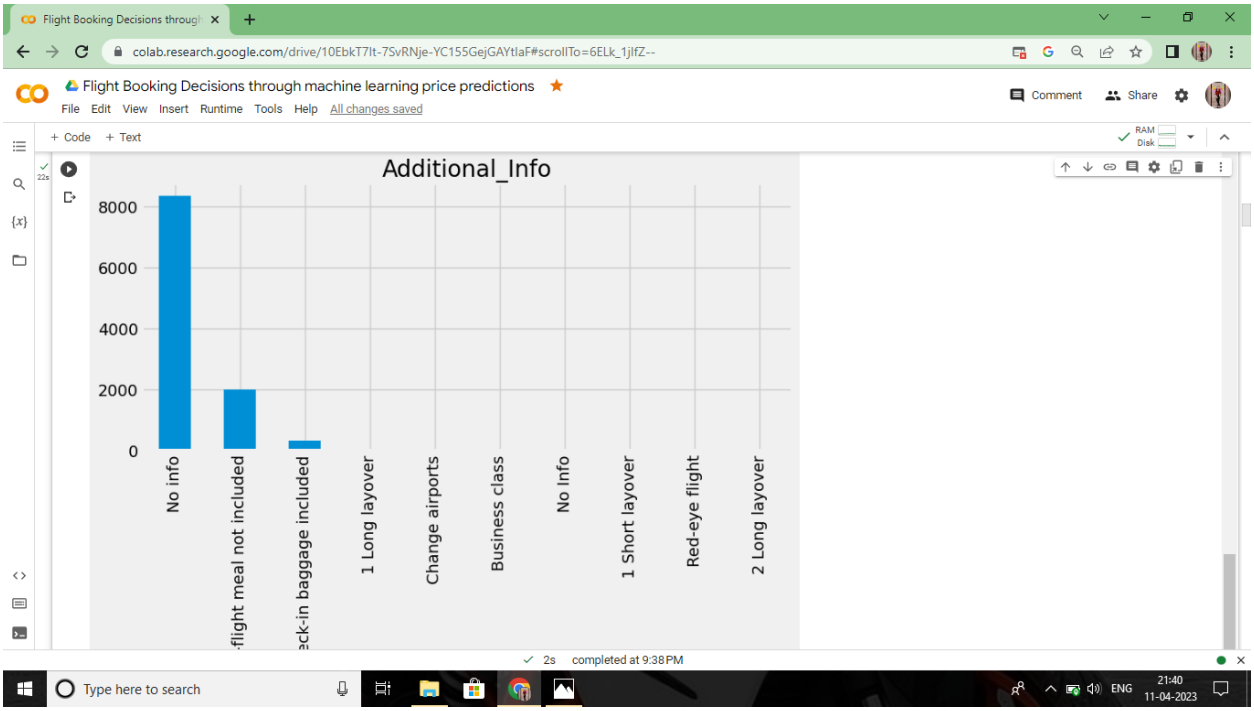
RAM Disk

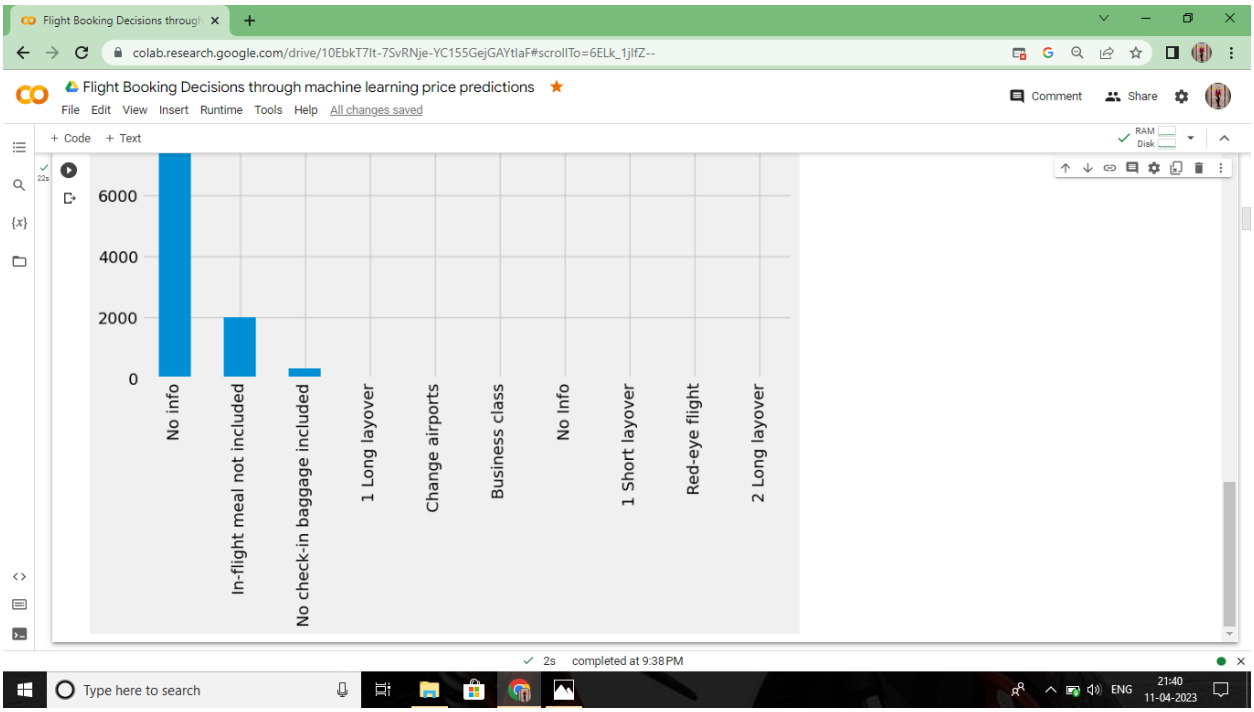


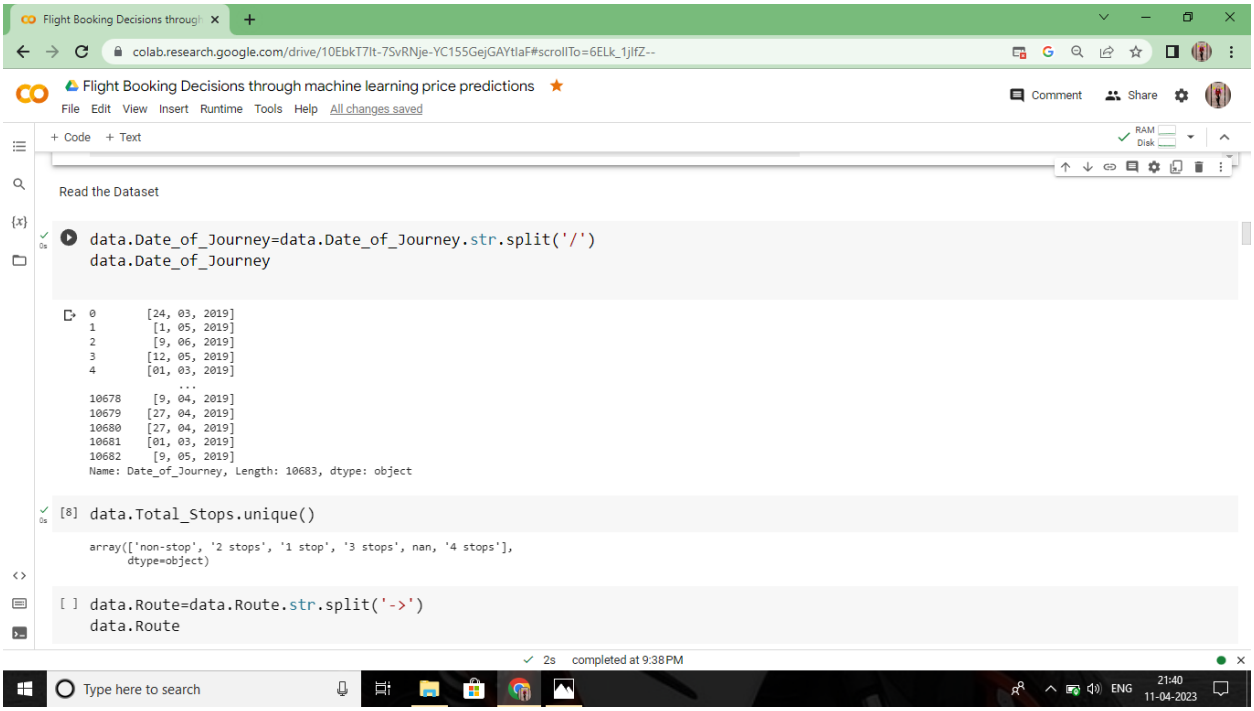
2s completed at 9:38 PM

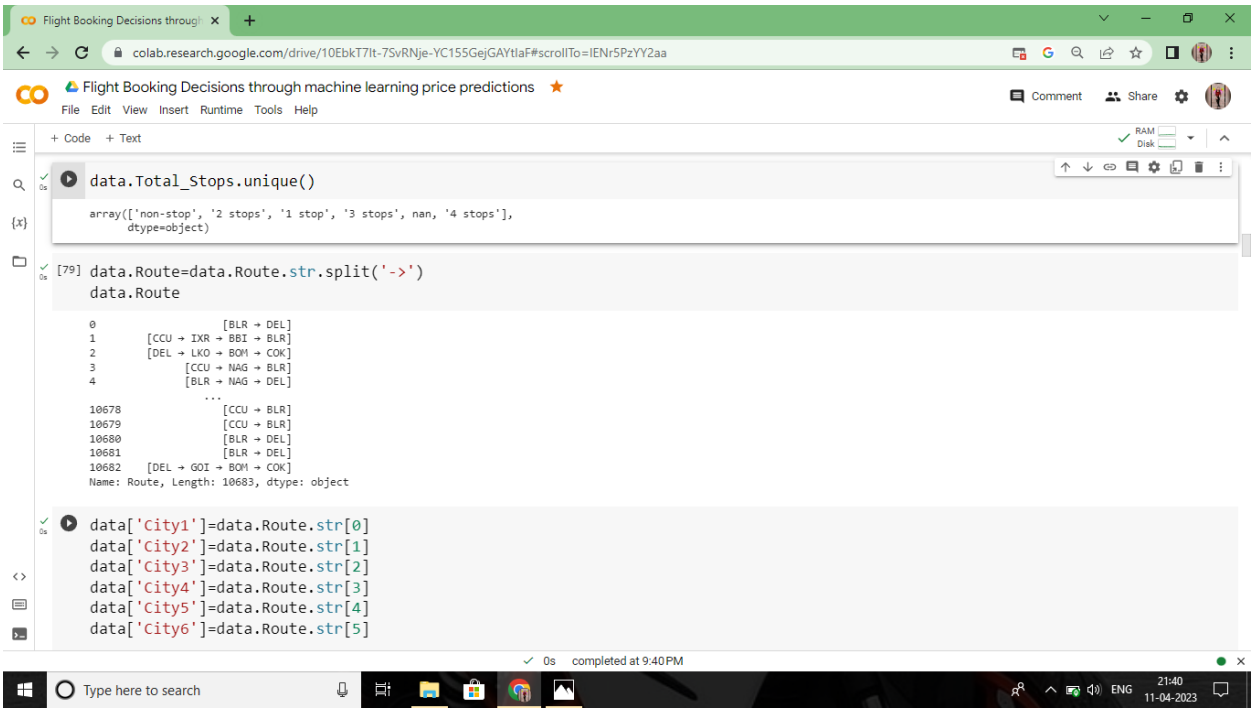
Type here to search

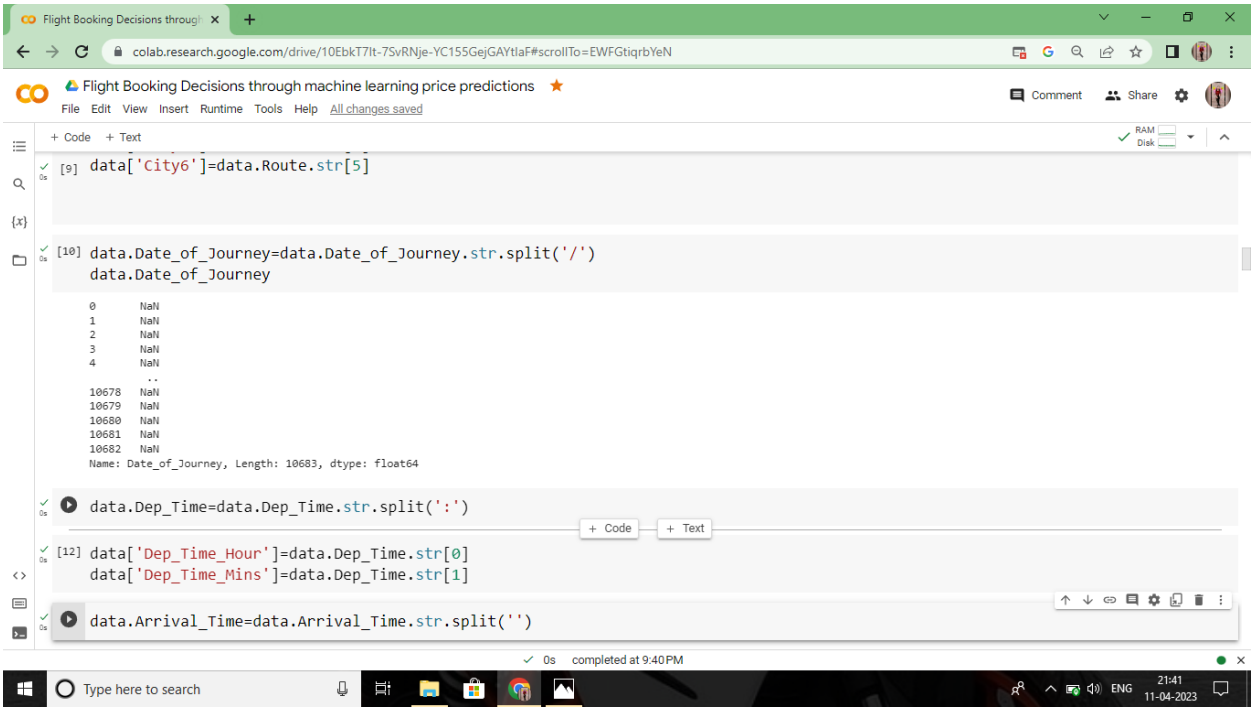
ENG 21:39 11-04-2023

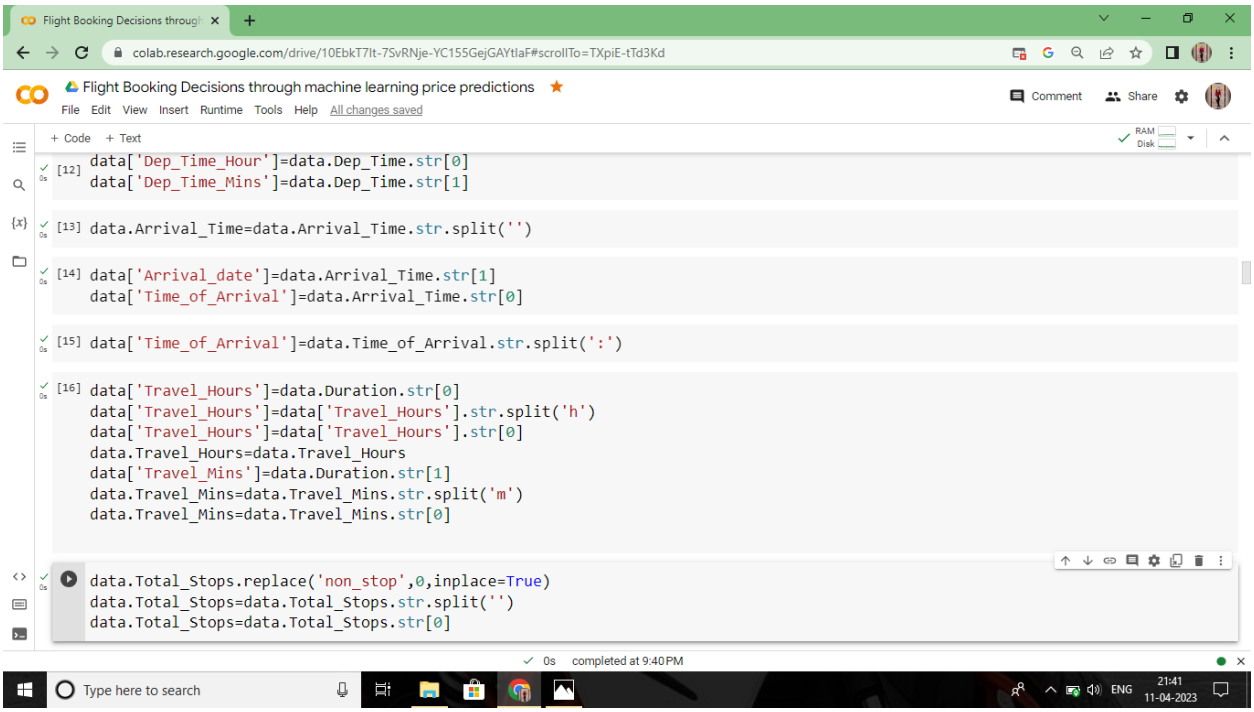

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=6ELk_1jlfZ--
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Additional_Info
8000
6000
4000
2000
0
No info
flight meal not included
ack-in baggage included
1 Long layover
Change airports
Business class
No Info
1 Short layover
Red-eye flight
2 Long layover
2s completed at 9:38 PM
Type here to search
ENG
21:40
11-04-2023

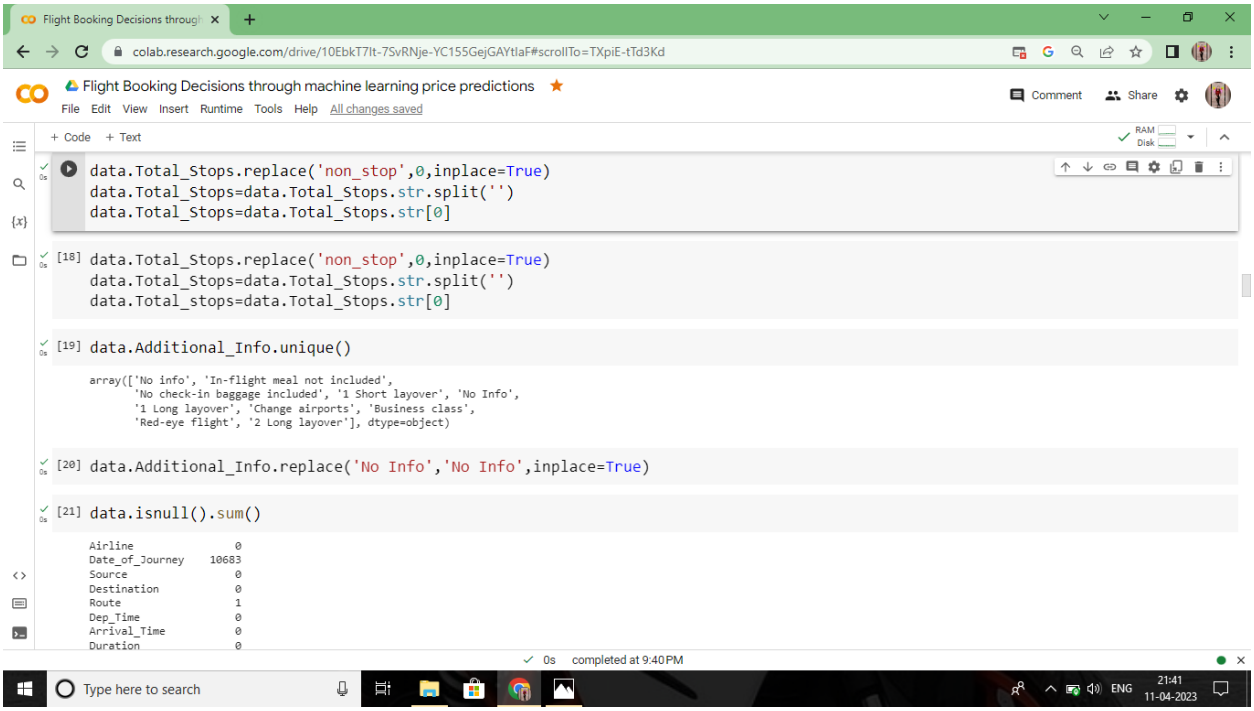

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=6ELk_1jlfZ--
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
6000
4000
2000
0
No info
In-flight meal not included
No check-in baggage included
1 Long layover
Change airports
Business class
No Info
1 Short layover
Red-eye flight
2 Long layover
2s completed at 9:38 PM
Type here to search
ENG
21:40
11-04-2023

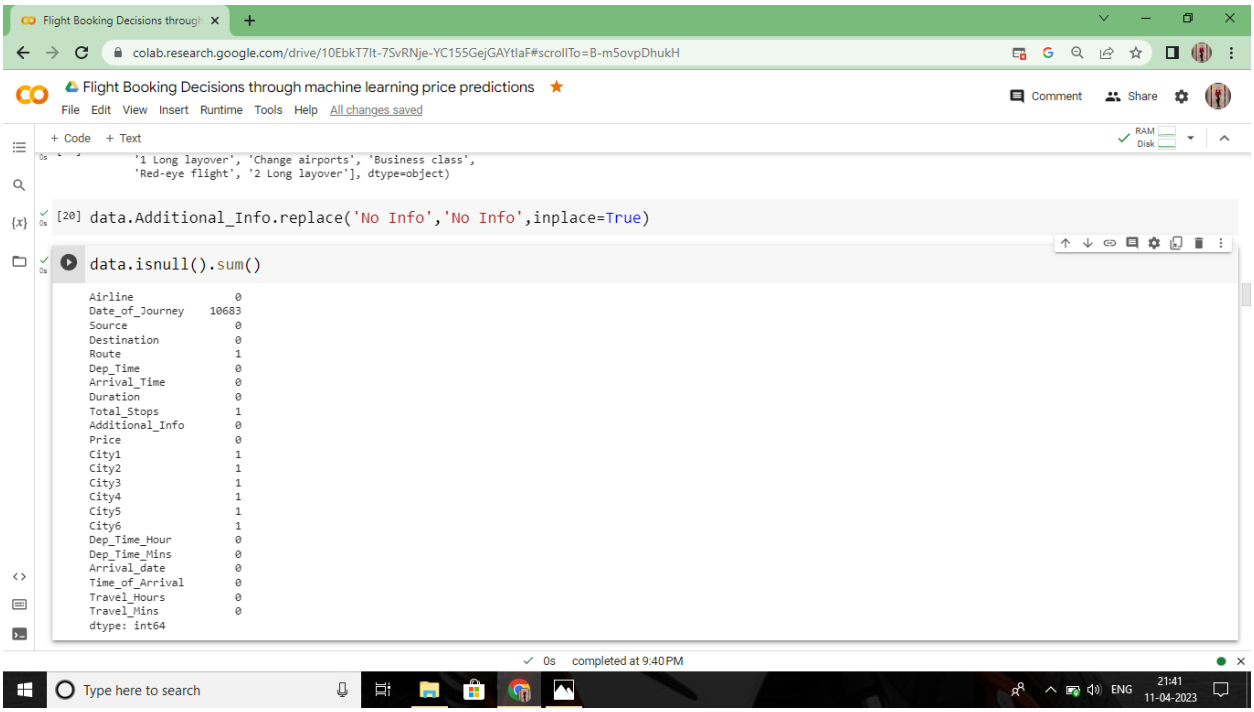

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=6ELk_1jlfZ--
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Read the Dataset
data.Date_of_Journey=data.Date_of_Journey.str.split('/')
data.Date_of_Journey
0 [24, 03, 2019]
1 [1, 05, 2019]
2 [9, 06, 2019]
3 [12, 05, 2019]
4 [01, 03, 2019]
...
10678 [9, 04, 2019]
10679 [27, 04, 2019]
10680 [27, 04, 2019]
10681 [01, 03, 2019]
10682 [9, 05, 2019]
Name: Date_of_Journey, Length: 10683, dtype: object
[8] data.Total_Stops.unique()
array(['non-stop', '2 stops', '1 stop', '3 stops', nan, '4 stops'],
dtype=object)
[ ] data.Route=data.Route.str.split('->')
data.Route
2s completed at 9:38PM
Type here to search
ENG
21:40
11-04-2023

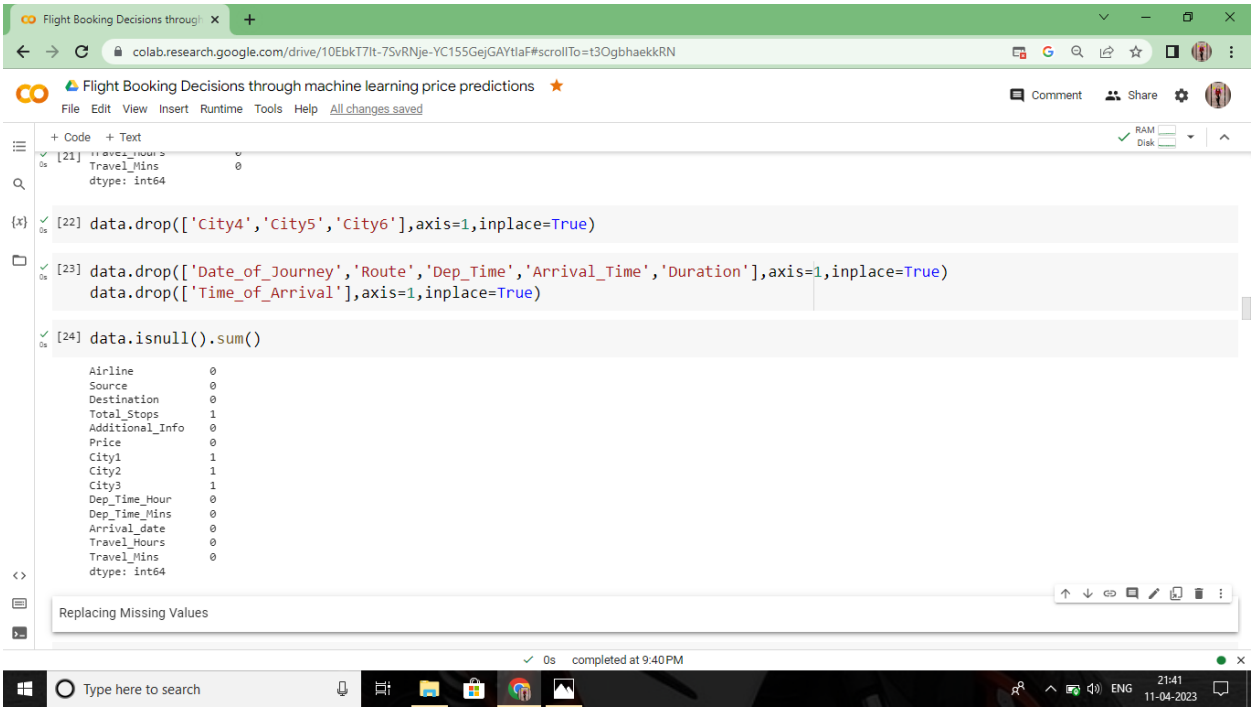

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=IENr5PzYY2aa
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
RAM
Disk
data.Total_Stops.unique()
array(['non-stop', '2 stops', '1 stop', '3 stops', nan, '4 stops'],
dtype=object)
[79] 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: 10683, dtype: object
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]
0s completed at 9:40PM
Type here to search
ENG
21:40
11-04-2023

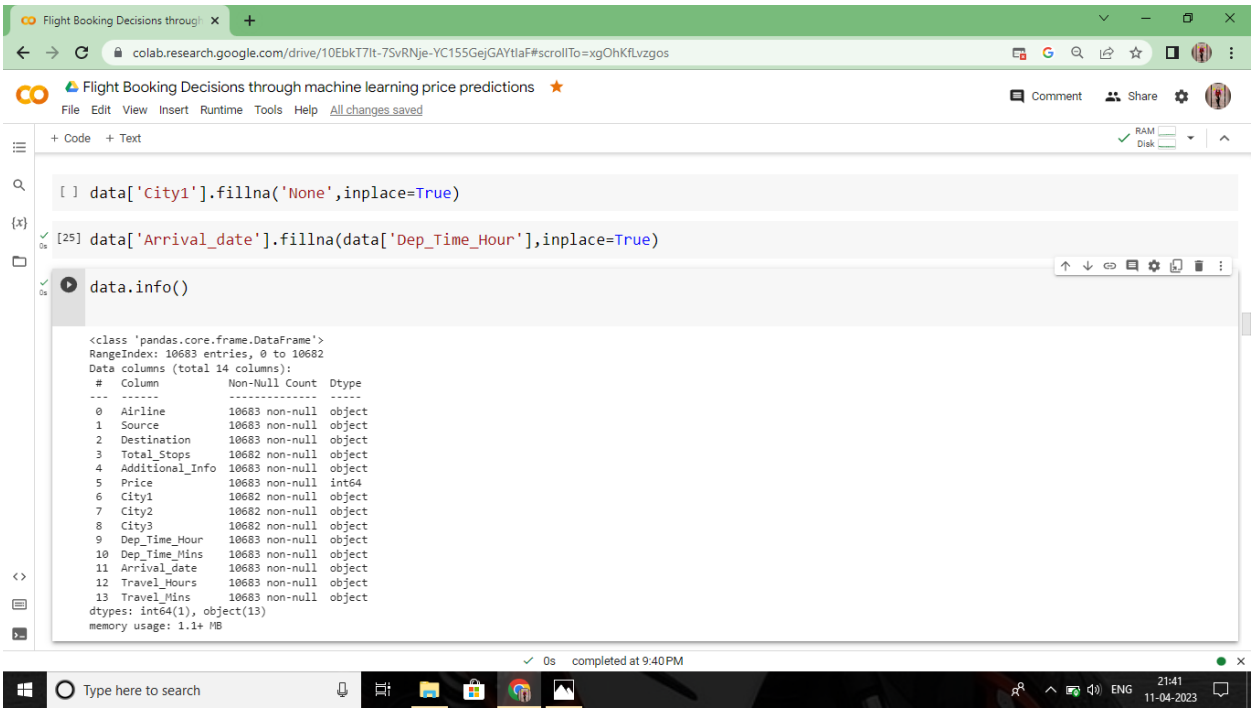

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=EWFGtiqrbYeN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[9] data['City6']=data.Route.str[5]
[10] data.Date_of_Journey=data.Date_of_Journey.str.split('/')
data.Date_of_Journey
0 NaN
1 NaN
2 NaN
3 NaN
4 NaN
10678 NaN
10679 NaN
10680 NaN
10681 NaN
10682 NaN
Name: Date_of_Journey, Length: 10683, dtype: float64
data.Dep_Time=data.Dep_Time.str.split(':')
+ Code + Text
[12] data['Dep_Time_Hour']=data.Dep_Time.str[0]
data['Dep_Time_Mins']=data.Dep_Time.str[1]
data.Arrival_Time=data.Arrival_Time.str.split('')
0s completed at 9:40PM
Type here to search
ENG
21:41
11-04-2023

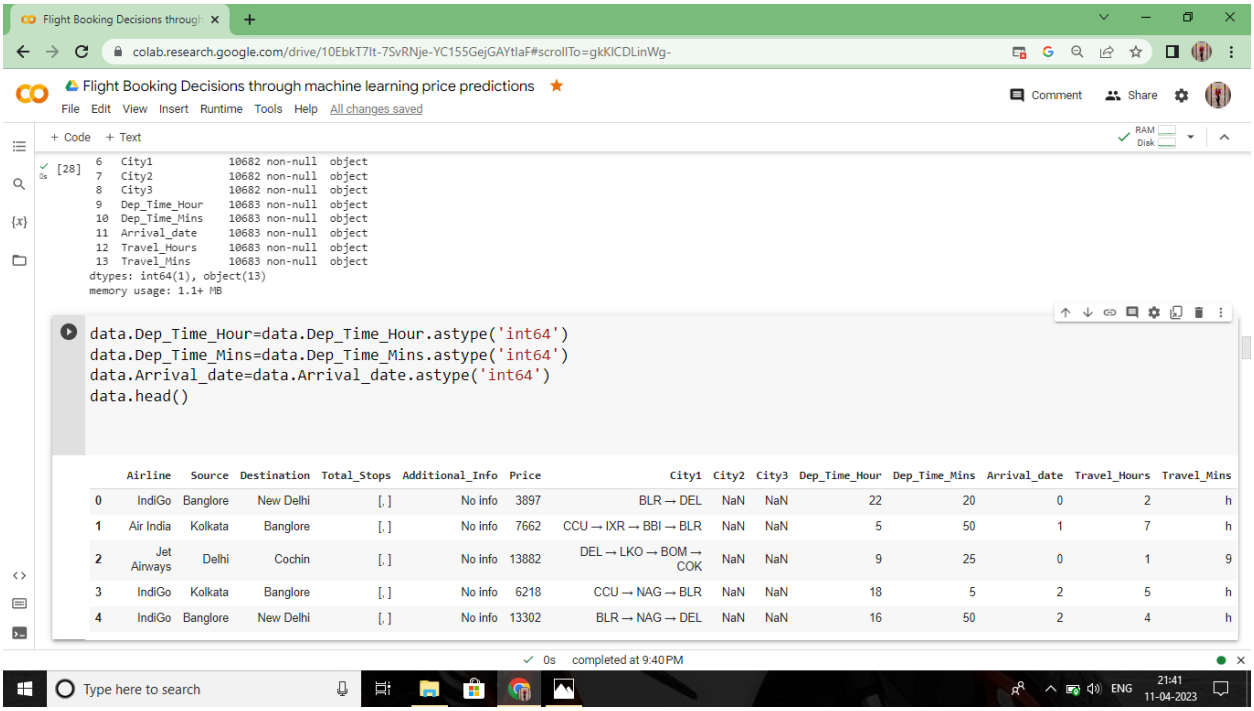

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=TXpiE-tTd3Kd
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[12] data['Dep_Time_Hour']=data.Dep_Time.str[0]
data['Dep_Time_Mins']=data.Dep_Time.str[1]
[13] data.Arrival_Time=data.Arrival_Time.str.split('')
[14] data['Arrival_date']=data.Arrival_Time.str[1]
data['Time_of_Arrival']=data.Arrival_Time.str[0]
[15] data['Time_of_Arrival']=data.Time_of_Arrival.str.split(':')
[16] 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]
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]
0s completed at 9:40PM
Type here to search
ENG
21:41
11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=TXpiE-tTd3Kd
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
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]
[18] 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]
[19] 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)
[20] data.Additional_Info.replace('No Info','No Info',inplace=True)
[21] data.isnull().sum()
Airline 0
Date_of_Journey 10683
Source 0
Destination 0
Route 1
Dep_Time 0
Arrival_Time 0
Duration 0
0s completed at 9:40PM
Type here to search
ENG
21:41
11-04-2023

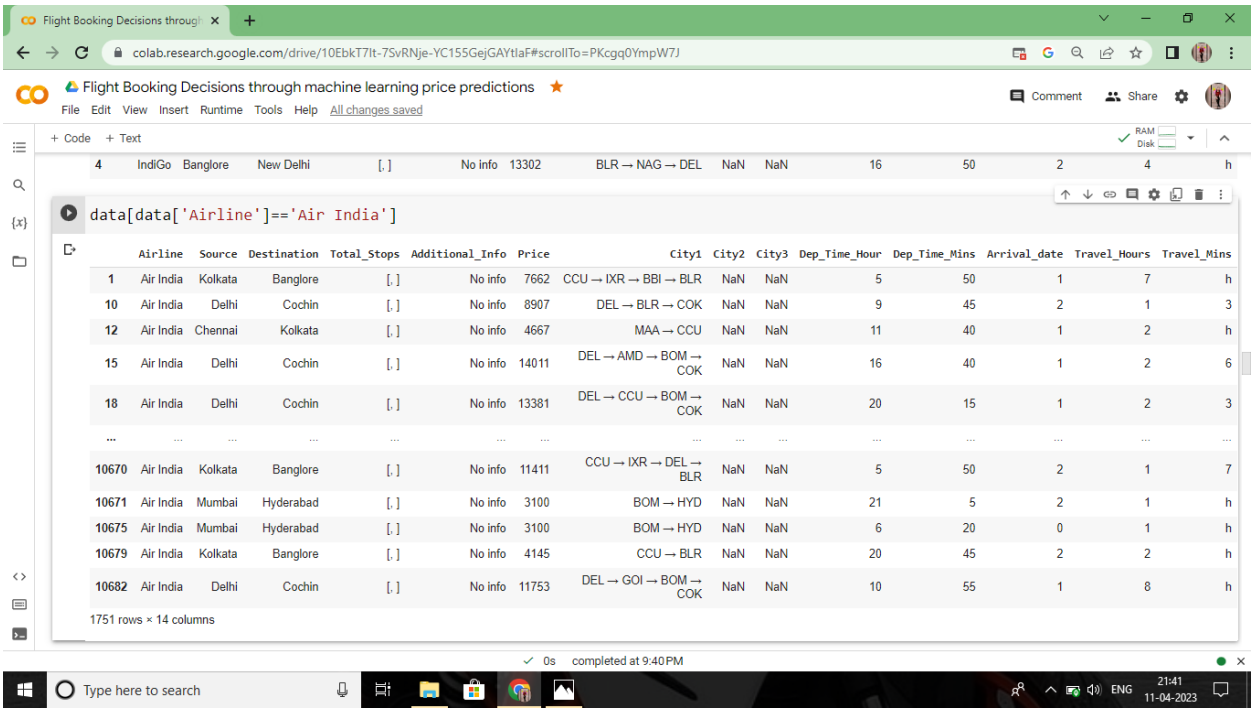

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=B-m5ovpDhukH
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
'1 Long layover', 'Change airports', 'Business class',
'Red-eye flight', '2 Long layover'], dtype=object)
[20] data.Additional_Info.replace('No Info','No Info',inplace=True)
data.isnull().sum()
Airline 0
Date_of_Journey 10683
Source 0
Destination 0
Route 1
Dep_Time 0
Arrival_Time 0
Duration 0
Total_Stops 1
Additional_Info 0
Price 0
City1 1
City2 1
City3 1
City4 1
City5 1
City6 1
Dep_Time_Hour 0
Dep_Time_Mins 0
Arrival_date 0
Time_of_Arrival 0
Travel_Hours 0
Travel_Mins 0
dtype: int64
0s completed at 9:40PM
Type here to search
ENG
21:41
11-04-2023


Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Travel_Mins 0
dtype: int64
[22] data.drop(['City4','City5','City6'],axis=1,inplace=True)
[23] data.drop(['Date_of_Journey','Route','Dep_Time','Arrival_Time','Duration'],axis=1,inplace=True)
data.drop(['Time_of_Arrival'],axis=1,inplace=True)
[24] data.isnull().sum()
Airline 0
Source 0
Destination 0
Total_Stops 1
Additional_Info 0
Price 0
City1 1
City2 1
City3 1
Dep_Time_Hour 0
Dep_Time_Mins 0
Arrival_date 0
Travel_Hours 0
Travel_Mins 0
dtype: int64
Replacing Missing Values
0s completed at 9:40 PM
Type here to search
ENG 21:41 11-04-2023

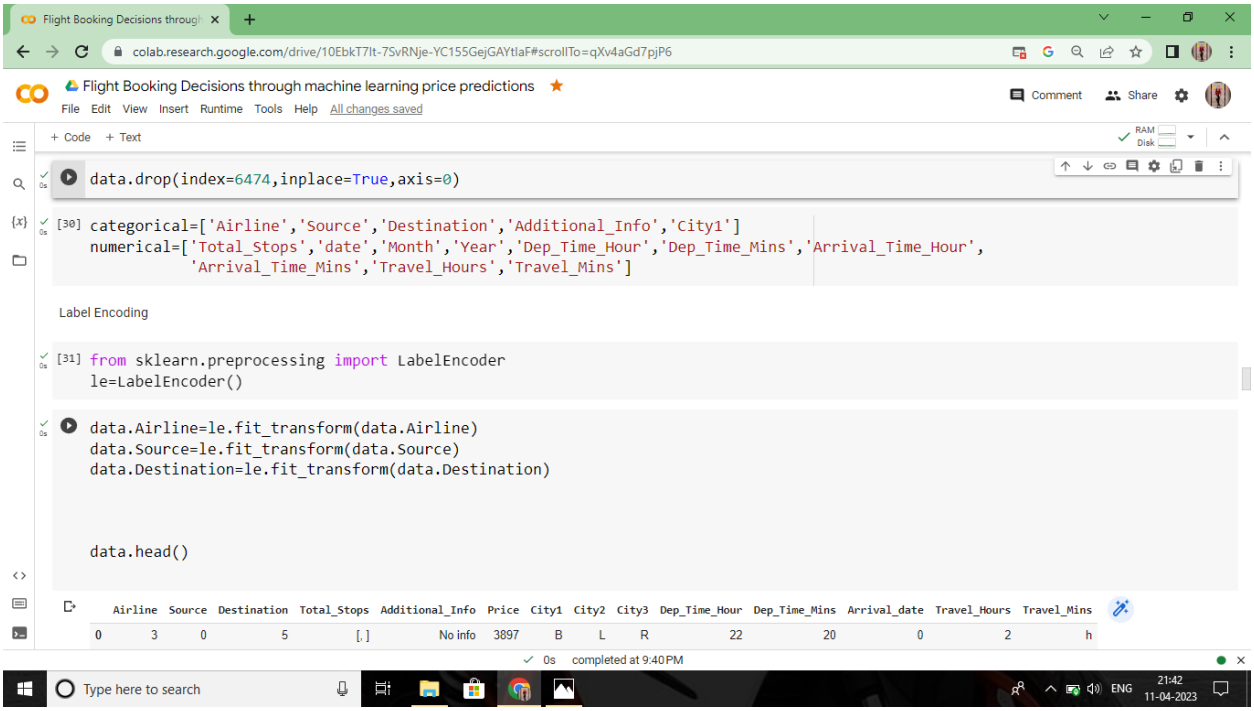

Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
[ ] data['City1'].fillna('None',inplace=True)
[25] data['Arrival_date'].fillna(data['Dep_Time_Hour'],inplace=True)
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10683 entries, 0 to 10682
Data columns (total 14 columns):
Column Non-Null Count Dtype
0 Airline 10683 non-null object
1 Source 10683 non-null object
2 Destination 10683 non-null object
3 Total_Stops 10682 non-null object
4 Additional_Info 10683 non-null object
5 Price 10683 non-null int64
6 City1 10682 non-null object
7 City2 10682 non-null object
8 City3 10682 non-null object
9 Dep_Time_Hour 10683 non-null object
10 Dep_Time_Mins 10683 non-null object
11 Arrival_date 10683 non-null object
12 Travel_Hours 10683 non-null object
13 Travel_Mins 10683 non-null object
dtypes: int64(1), object(13)
memory usage: 1.1+ MB
0s completed at 9:40 PM
Type here to search
ENG 21:41 11-04-2023

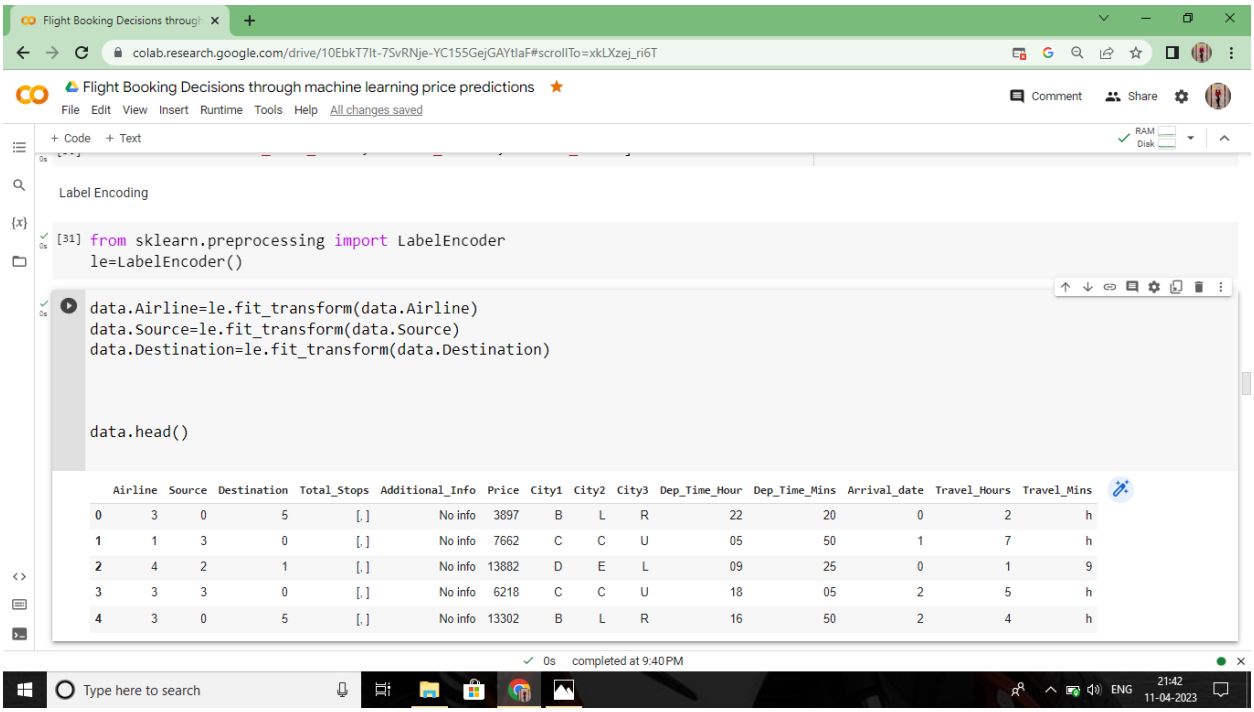


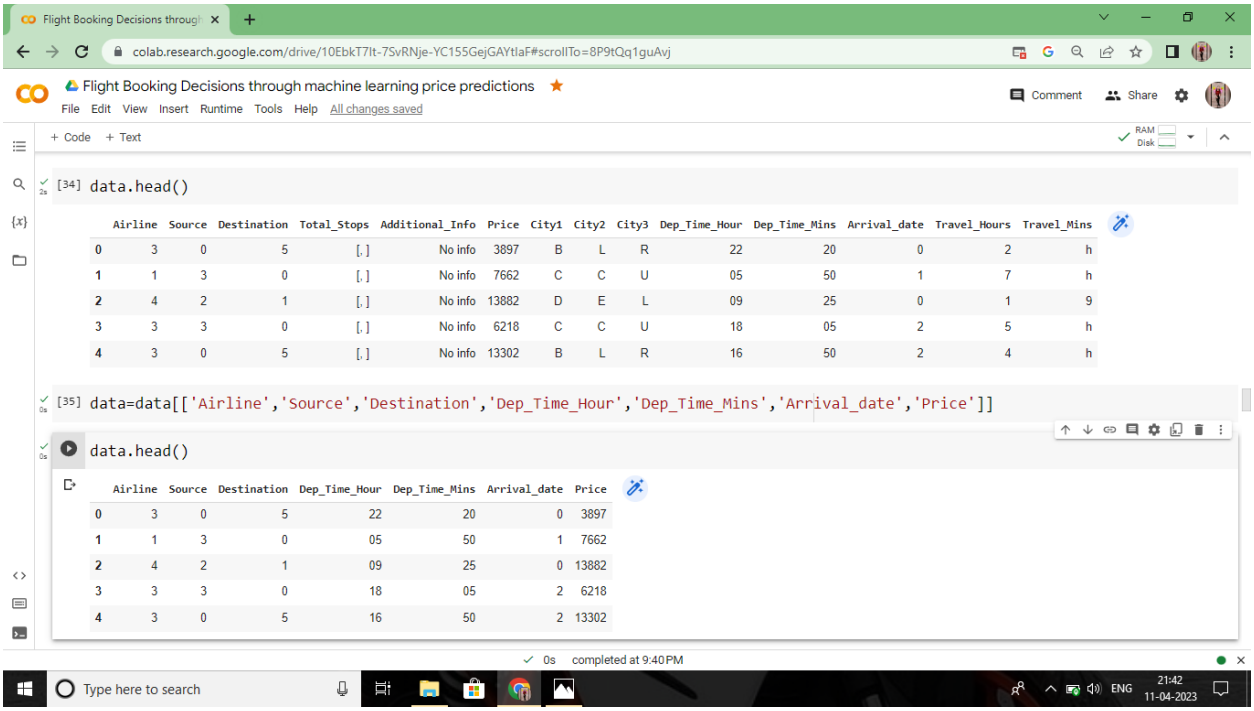
	Airline	Source	Destination	Total_Stops	Additional_Info	Price	City1	City2	City3	Dep_Time_Hour	Dep_Time_Mins	Arrival_date	Travel_Hours	Travel_Mins
0	IndiGo	Banglore	New Delhi	[,]	No info	3897	BLR → DEL	NaN	NaN	22	20	0	2	h
1	Air India	Kolkata	Banglore	[,]	No info	7662	CCU → IXR → BBI → BLR	NaN	NaN	5	50	1	7	h
2	Jet Airways	Delhi	Cochin	[,]	No info	13882	DEL → LKO → BOM → COK	NaN	NaN	9	25	0	1	9
3	IndiGo	Kolkata	Banglore	[,]	No info	6218	CCU → NAG → BLR	NaN	NaN	18	5	2	5	h
4	IndiGo	Banglore	New Delhi	[,]	No info	13302	BLR → NAG → DEL	NaN	NaN	16	50	2	4	h



	Airline	Source	Destination	Total_Stops	Additional_Info	Price	City1	City2	City3	Dep_Time_Hour	Dep_Time_Mins	Arrival_date	Travel_Hours	Travel_Mins
1	Air India	Kolkata	Banglore	[,]	No info	7662	CCU → IXR → BBI → BLR	NaN	NaN	5	50	1	7	h
10	Air India	Delhi	Cochin	[,]	No info	8907	DEL → BLR → COK	NaN	NaN	9	45	2	1	3
12	Air India	Chennai	Kolkata	[,]	No info	4667	MAA → CCU	NaN	NaN	11	40	1	2	h
15	Air India	Delhi	Cochin	[,]	No info	14011	DEL → AMD → BOM → COK	NaN	NaN	16	40	1	2	6
18	Air India	Delhi	Cochin	[,]	No info	13381	DEL → CCU → BOM → COK	NaN	NaN	20	15	1	2	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10670	Air India	Kolkata	Banglore	[,]	No info	11411	CCU → IXR → DEL → BLR	NaN	NaN	5	50	2	1	7
10671	Air India	Mumbai	Hyderabad	[,]	No info	3100	BOM → HYD	NaN	NaN	21	5	2	1	h
10675	Air India	Mumbai	Hyderabad	[,]	No info	3100	BOM → HYD	NaN	NaN	6	20	0	1	h
10679	Air India	Kolkata	Banglore	[,]	No info	4145	CCU → BLR	NaN	NaN	20	45	2	2	h
10682	Air India	Delhi	Cochin	[,]	No info	11753	DEL → GOI → BOM → COK	NaN	NaN	10	55	1	8	h

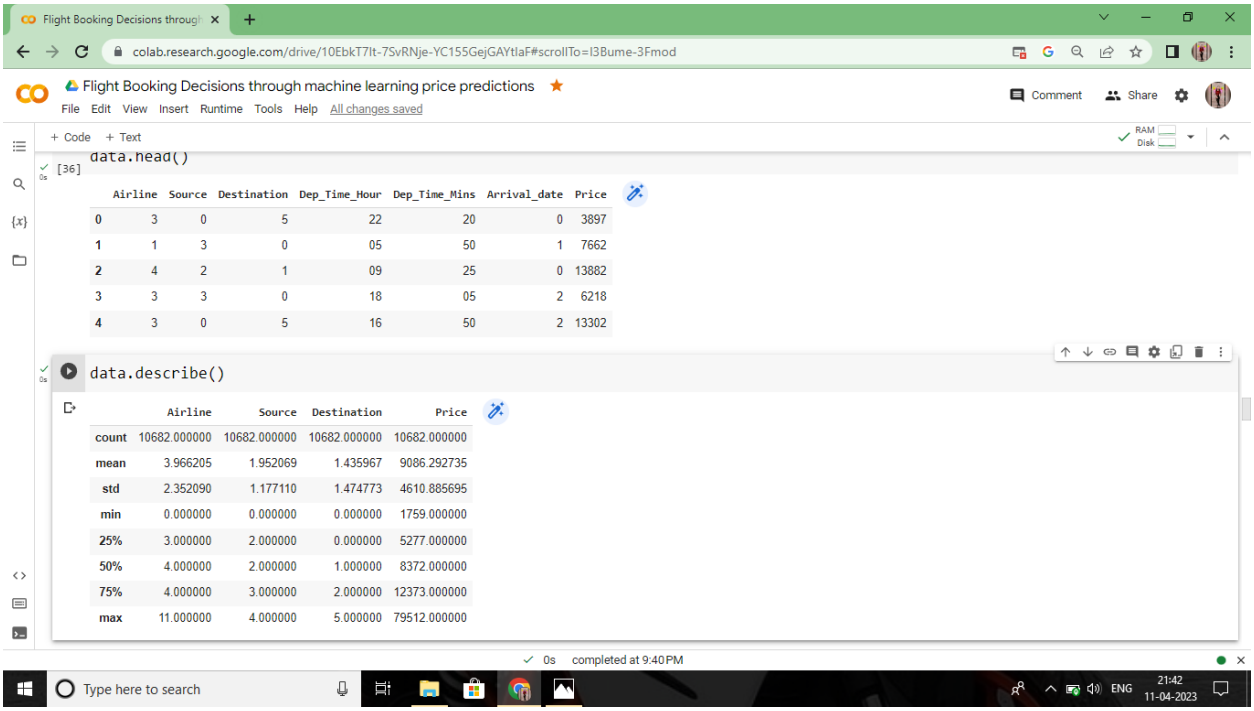

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qXv4aGd7pjP6
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
data.drop(index=6474,inplace=True,axis=0)
[30] categorical=['Airline','Source','Destination','Additional_Info','City1']
numerical=['Total_Stops','date','Month','Year','Dep_Time_Hour','Dep_Time_Mins','Arrival_Time_Hour',
'Arrival_Time_Mins','Travel_Hours','Travel_Mins']
Label Encoding
[31] from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data.Airline=le.fit_transform(data.Airline)
data.Source=le.fit_transform(data.Source)
data.Destination=le.fit_transform(data.Destination)
data.head()
Airline Source Destination Total_Stops Additional_Info Price City1 City2 City3 Dep_Time_Hour Dep_Time_Mins Arrival_date Travel_Hours Travel_Mins
0 3 0 5 [.] No info 3897 B L R 22 20 0 2 h
0s completed at 9:40 PM
Type here to search
ENG
21:42
11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=xkLXzej_ri6T
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Label Encoding
[31] from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data.Airline=le.fit_transform(data.Airline)
data.Source=le.fit_transform(data.Source)
data.Destination=le.fit_transform(data.Destination)
data.head()
Airline Source Destination Total_Stops Additional_Info Price City1 City2 City3 Dep_Time_Hour Dep_Time_Mins Arrival_date Travel_Hours Travel_Mins
0 3 0 5 [.] No info 3897 B L R 22 20 0 2 h
1 1 3 0 [.] No info 7662 C C U 05 50 1 7 h
2 4 2 1 [.] No info 13882 D E L 09 25 0 1 9
3 3 3 0 [.] No info 6218 C C U 18 05 2 5 h
4 3 0 5 [.] No info 13302 B L R 16 50 2 4 h
0s completed at 9:40 PM
Type here to search
ENG
21:42
11-04-2023



	Airline	Source	Destination	Total_Stops	Additional_Info	Price	City1	City2	City3	Dep_Time_Hour	Dep_Time_Mins	Arrival_date	Travel_Hours	Travel_Mins
0	3	0	5	[.]	No info	3897	B	L	R	22	20	0	2	h
1	1	3	0	[.]	No info	7662	C	C	U	05	50	1	7	h
2	4	2	1	[.]	No info	13882	D	E	L	09	25	0	1	9
3	3	3	0	[.]	No info	6218	C	C	U	18	05	2	5	h
4	3	0	5	[.]	No info	13302	B	L	R	16	50	2	4	h

	Airline	Source	Destination	Dep_Time_Hour	Dep_Time_Mins	Arrival_date	Price
0	3	0	5	22	20	0	3897
1	1	3	0	05	50	1	7662
2	4	2	1	09	25	0	13882
3	3	3	0	18	05	2	6218
4	3	0	5	16	50	2	13302



	Airline	Source	Destination	Dep_Time_Hour	Dep_Time_Mins	Arrival_date	Price
0	3	0	5	22	20	0	3897
1	1	3	0	05	50	1	7662
2	4	2	1	09	25	0	13882
3	3	3	0	18	05	2	6218
4	3	0	5	16	50	2	13302

	Airline	Source	Destination	Price
count	10682.000000	10682.000000	10682.000000	10682.000000
mean	3.966205	1.952069	1.435967	9086.292735
std	2.352090	1.177110	1.474773	4610.885695
min	0.000000	0.000000	0.000000	1759.000000
25%	3.000000	2.000000	0.000000	5277.000000
50%	4.000000	2.000000	1.000000	8372.000000
75%	4.000000	3.000000	2.000000	12373.000000
max	11.000000	4.000000	5.000000	79512.000000

Flight Booking Decisions through machine learning price predictions

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=ye2U8KCony2y

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

[36]

0	3	0	5	22	20	0	3897
1	1	3	0	05	50	1	7662
2	4	2	1	09	25	0	13882
3	3	3	0	18	05	2	6218
4	3	0	5	16	50	2	13302

```
data.describe()
```

	Airline	Source	Destination	Price
count	10682.000000	10682.000000	10682.000000	10682.000000
mean	3.966205	1.952069	1.435967	9086.292735
std	2.352090	1.177110	1.474773	4610.885695
min	0.000000	0.000000	0.000000	1759.000000
25%	3.000000	2.000000	0.000000	5277.000000
50%	4.000000	2.000000	1.000000	8372.000000
75%	4.000000	3.000000	2.000000	12373.000000
max	11.000000	4.000000	5.000000	79512.000000

Milestone 3:Explortory Data Analysis

0s completed at 9:40PM

Type here to search

ENG 21:42 11-04-2023

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=ySQay-3OVc8V

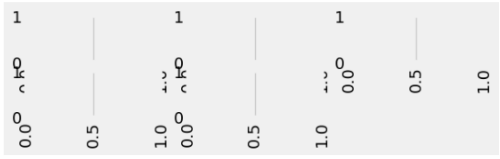
[38]

```
import seaborn as sns
c=1
plt.figure(figsize=(20,45))
```

```
<Figure size 2000x4500 with 0 Axes>
<Figure size 2000x4500 with 0 Axes>
```

```
for i in categorical:
  plt.subplot(6,3,c)

  plt.xticks(rotation=90)
  plt.tight_layout(pad=3.0)
  c=c+1
```



0s completed at 9:40PM

Type here to search

ENG 21:42 11-04-2023

Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=daUFDJH9Y7CH

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help

Comment Share

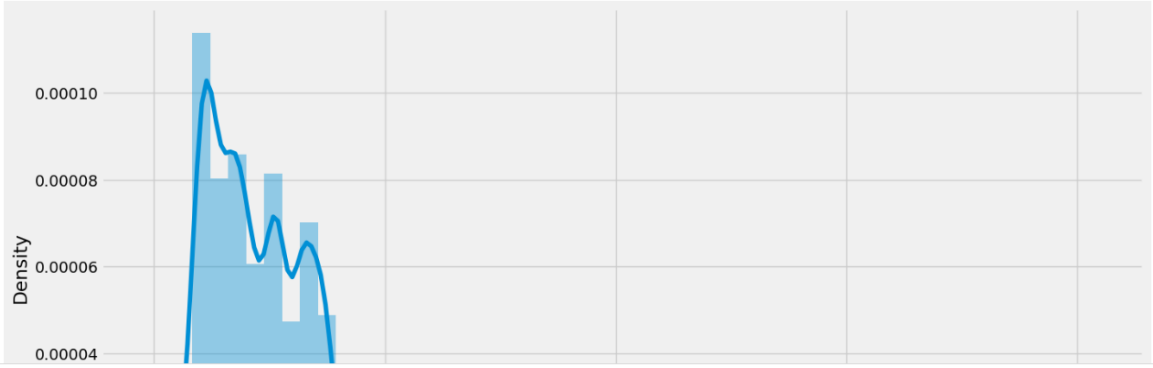
+ Code + Text

RAM Disk

```
plt.show()
```

```
[40] plt.figure(figsize=(15,8))
     sns.distplot(data.Price)
```

```
<Axes: xlabel='Price', ylabel='Density'>
```



0s completed at 9:42PM

Type here to search

ENG 21:42 11-04-2023

Flight Booking Decisions through

colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN

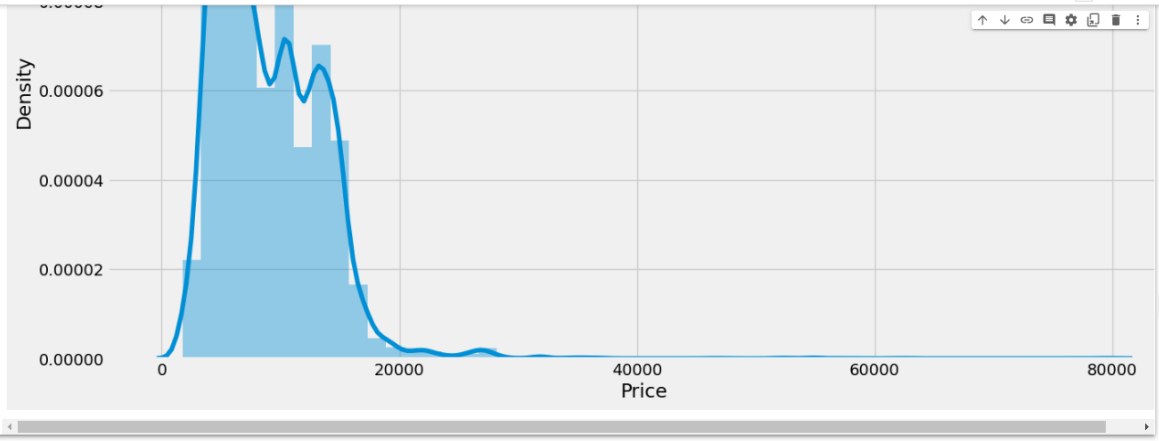
Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk




```
[41] sns.heatmap(data.corr(),annot=True)
```

```
<Axes: >
```

0s completed at 9:42PM

Type here to search

ENG 21:43 11-04-2023

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

```
sns.heatmap(data.corr(),annot=True)
```

<Axes: >


```
[42] # Detecting the Outliers
```

0s completed at 9:42PM

Type here to search

ENG 21:43 11-04-2023

Flight Booking Decisions through machine learning price predictions

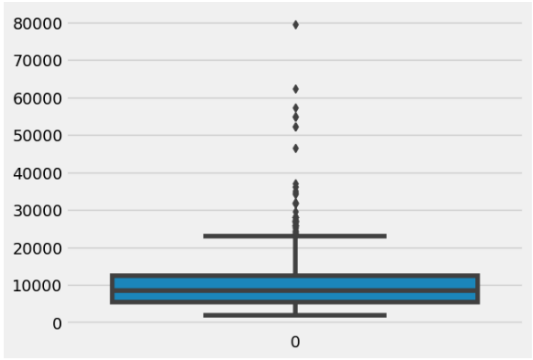
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

```
# Detecting the Outliers
import seaborn as sns
sns.boxplot(data['Price'])
```

<Axes: >

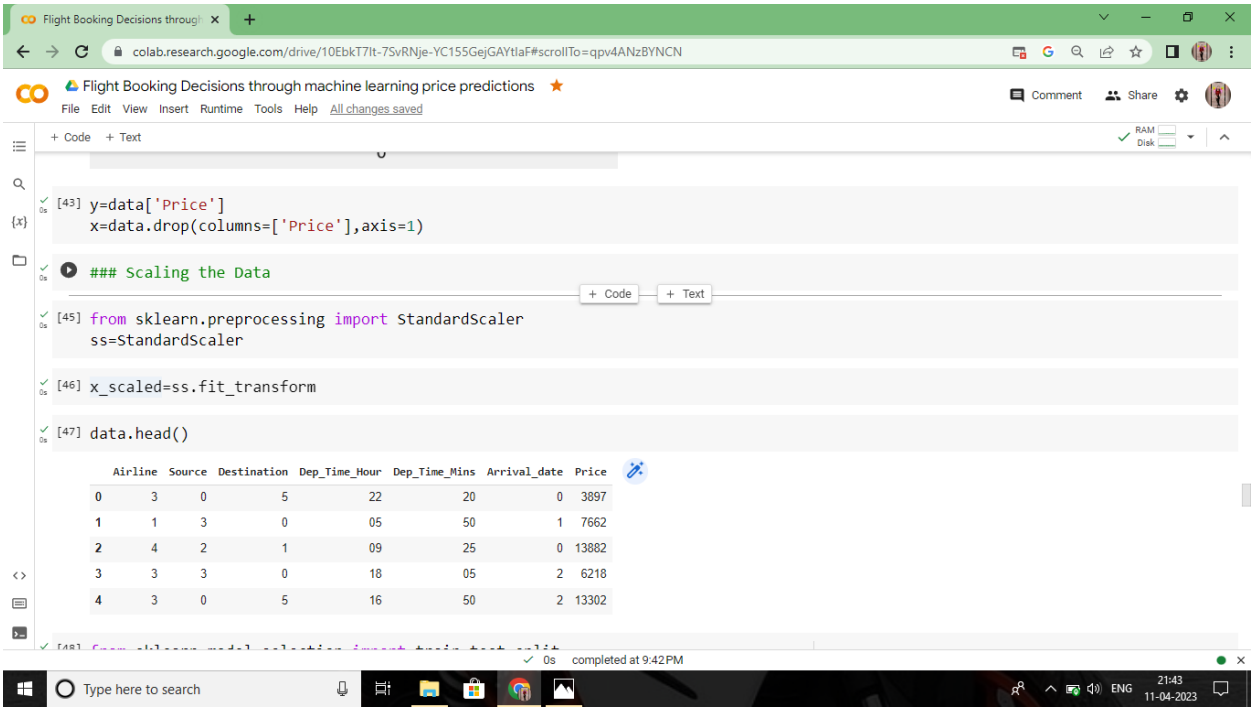


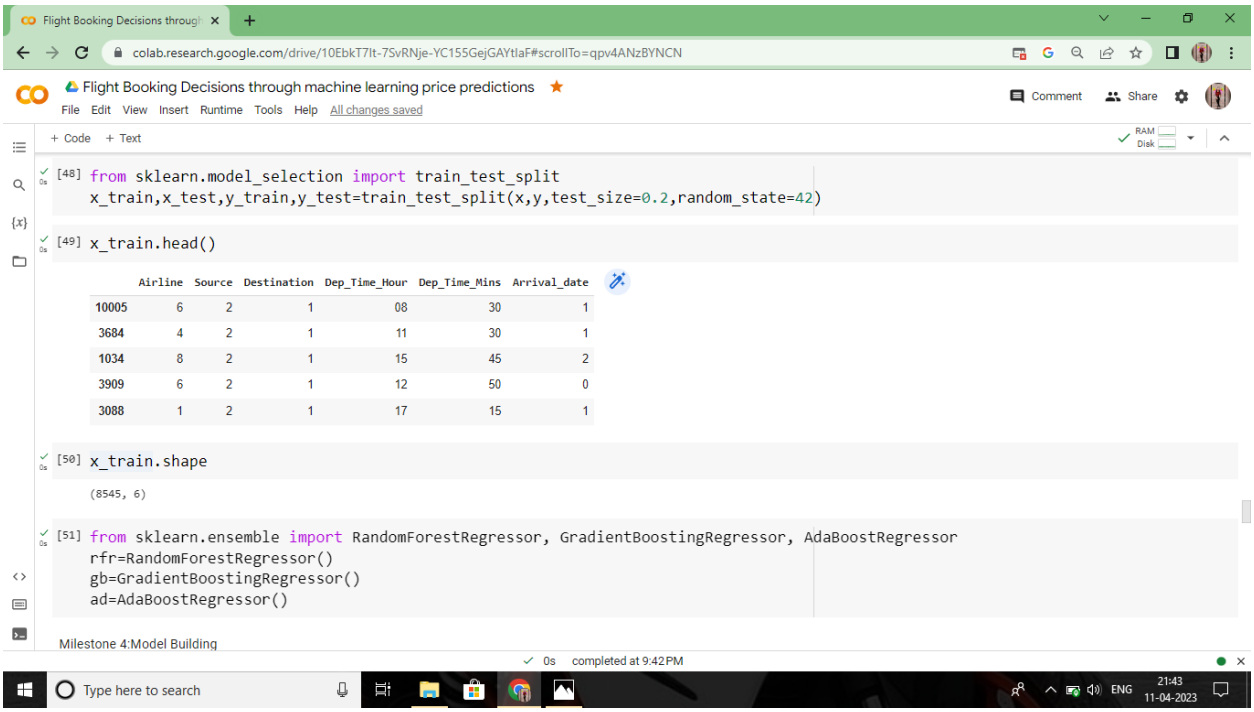

```
[43] y=data['Price']
```

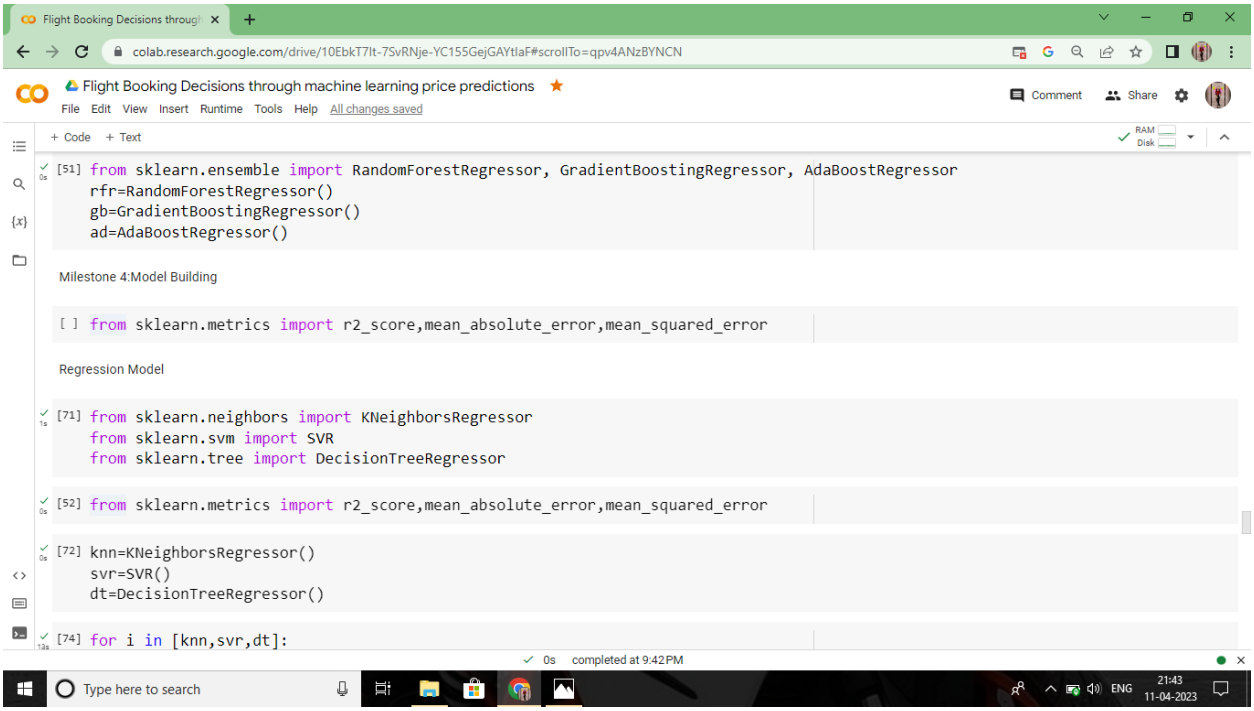
0s completed at 9:42PM

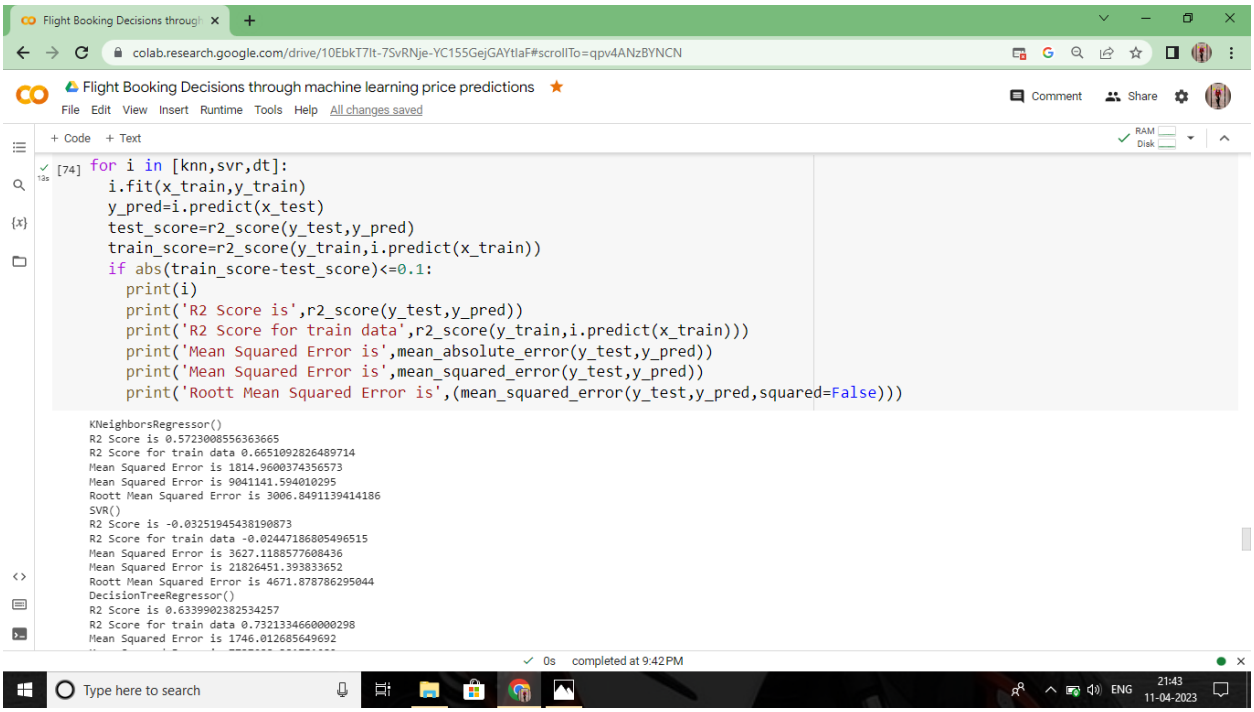
Type here to search

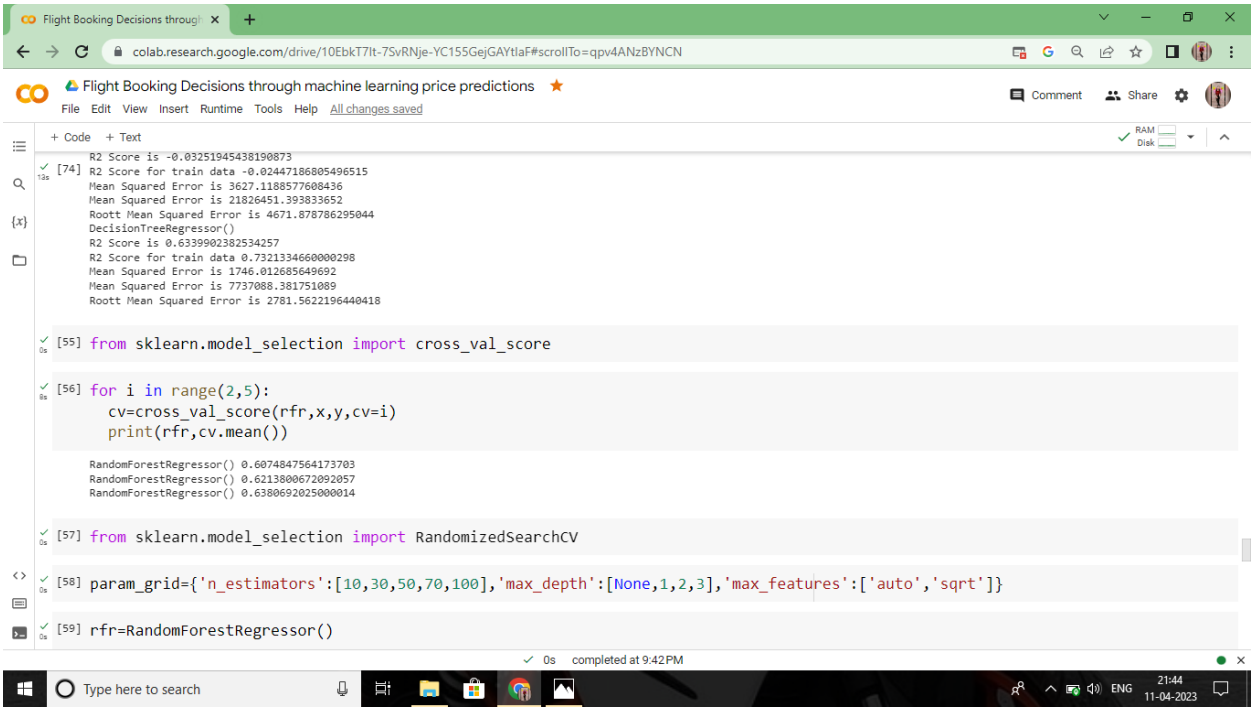
ENG 21:43 11-04-2023

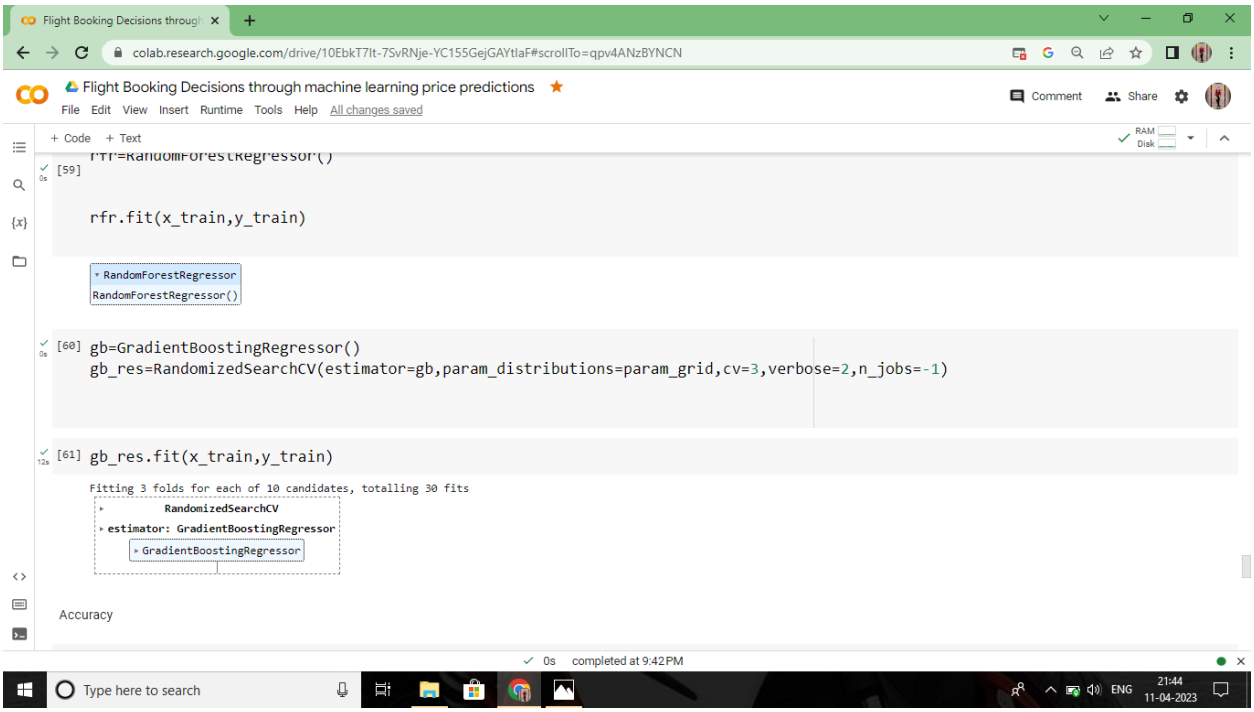

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[43] y=data['Price']
x=data.drop(columns=['Price'],axis=1)
Scaling the Data
+ Code + Text
[45] from sklearn.preprocessing import StandardScaler
ss=StandardScaler
[46] x_scaled=ss.fit_transform
[47] data.head()
Airline Source Destination Dep_Time_Hour Dep_Time_Mins Arrival_date Price
0 3 0 5 22 20 0 3897
1 1 3 0 05 50 1 7662
2 4 2 1 09 25 0 13882
3 3 3 0 18 05 2 6218
4 3 0 5 16 50 2 13302
0s completed at 9:42PM
Type here to search
ENG
21:43
11-04-2023

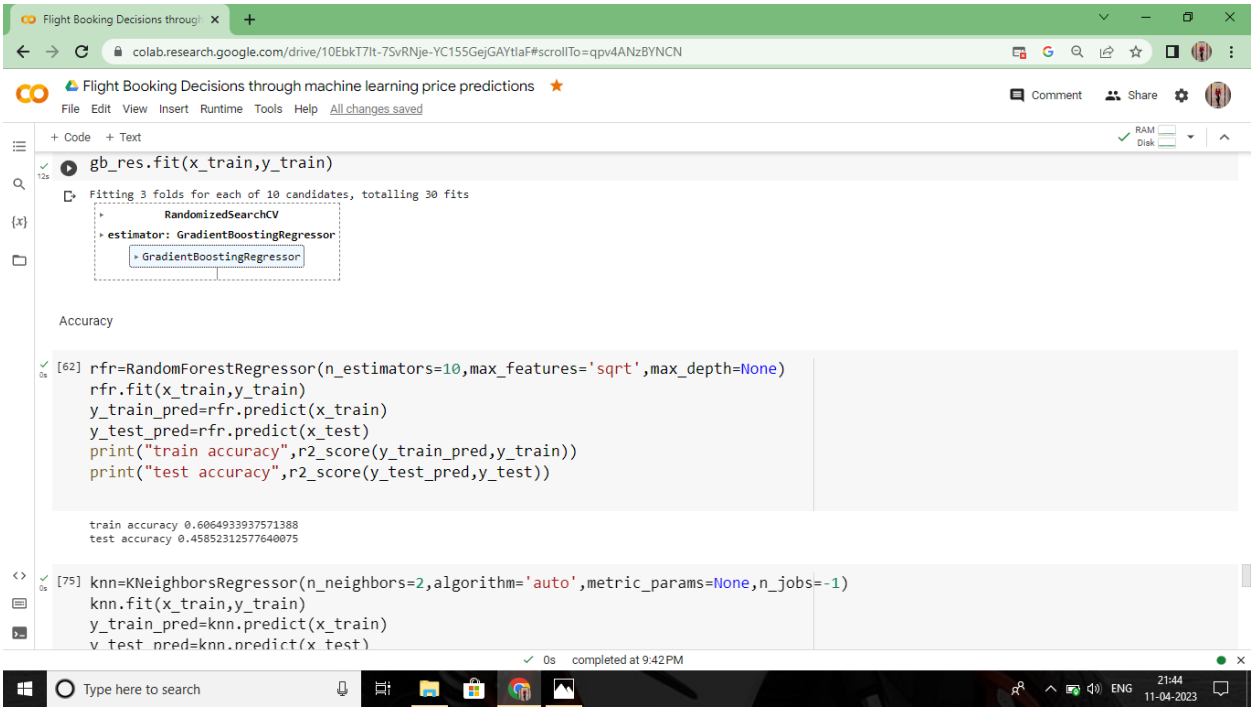

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[48] from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
[49] x_train.head()
Airline Source Destination Dep_Time_Hour Dep_Time_Mins Arrival_date
10005 6 2 1 08 30 1
3684 4 2 1 11 30 1
1034 8 2 1 15 45 2
3909 6 2 1 12 50 0
3088 1 2 1 17 15 1
[50] x_train.shape
(8545, 6)
[51] from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, AdaBoostRegressor
rfr=RandomForestRegressor()
gb=GradientBoostingRegressor()
ad=AdaBoostRegressor()
Milestone 4:Model Building
0s completed at 9:42PM
Type here to search
ENG
21:43
11-04-2023

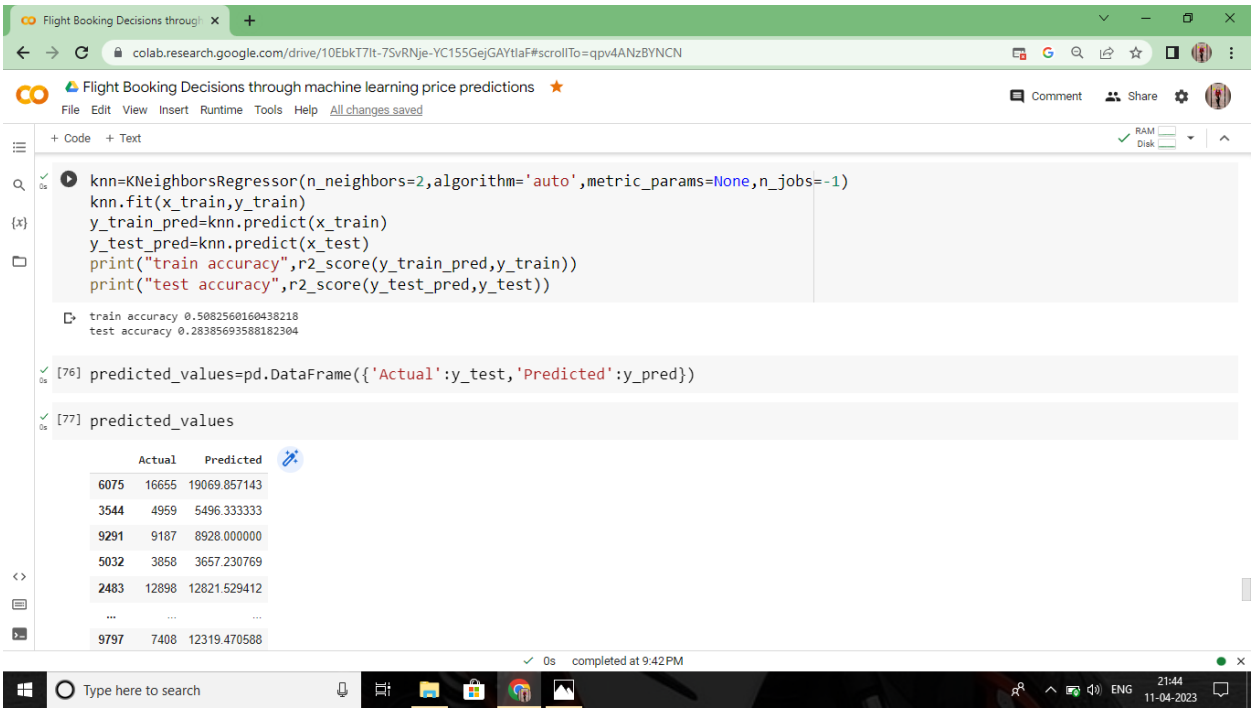

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[51] from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, AdaBoostRegressor
rfr=RandomForestRegressor()
gb=GradientBoostingRegressor()
ad=AdaBoostRegressor()
Milestone 4:Model Building
[ ] from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
Regression Model
[71] from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
from sklearn.tree import DecisionTreeRegressor
[52] from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
[72] knn=KNeighborsRegressor()
svr=SVR()
dt=DecisionTreeRegressor()
[74] for i in [knn,svr,dt]:
0s completed at 9:42PM
Type here to search
ENG
21:43
11-04-2023

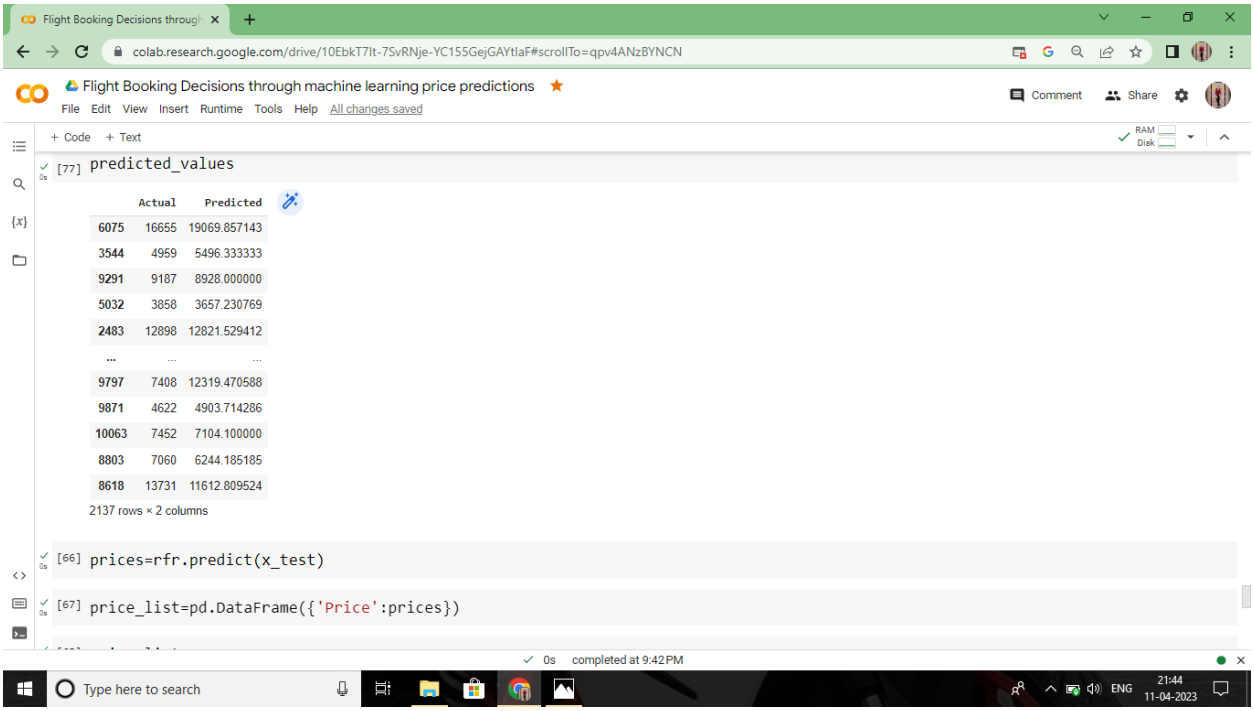

Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[74] for i in [knn,svr,dt]:
i.fit(x_train,y_train)
y_pred=i.predict(x_test)
test_score=r2_score(y_test,y_pred)
train_score=r2_score(y_train,i.predict(x_train))
if abs(train_score-test_score)<=0.1:
print(i)
print('R2 Score is',r2_score(y_test,y_pred))
print('R2 Score for train data',r2_score(y_train,i.predict(x_train)))
print('Mean Squared Error is',mean_absolute_error(y_test,y_pred))
print('Mean Squared Error is',mean_squared_error(y_test,y_pred))
print('Rootт Mean Squared Error is',(mean_squared_error(y_test,y_pred,squared=False)))
KNeighborsRegressor()
R2 Score is 0.5723008556363665
R2 Score for train data 0.6651092826489714
Mean Squared Error is 1814.9600374356573
Mean Squared Error is 9041141.594010295
Rootт Mean Squared Error is 3006.8491139414186
SVR()
R2 Score is -0.03251945438190873
R2 Score for train data -0.02447186805496515
Mean Squared Error is 3627.1188577608436
Mean Squared Error is 21826451.393833652
Rootт Mean Squared Error is 4671.878786295044
DecisionTreeRegressor()
R2 Score is 0.6339902382534257
R2 Score for train data 0.7321334660000298
Mean Squared Error is 1746.012685649692
0s completed at 9:42PM
Type here to search
ENG
21:43
11-04-2023

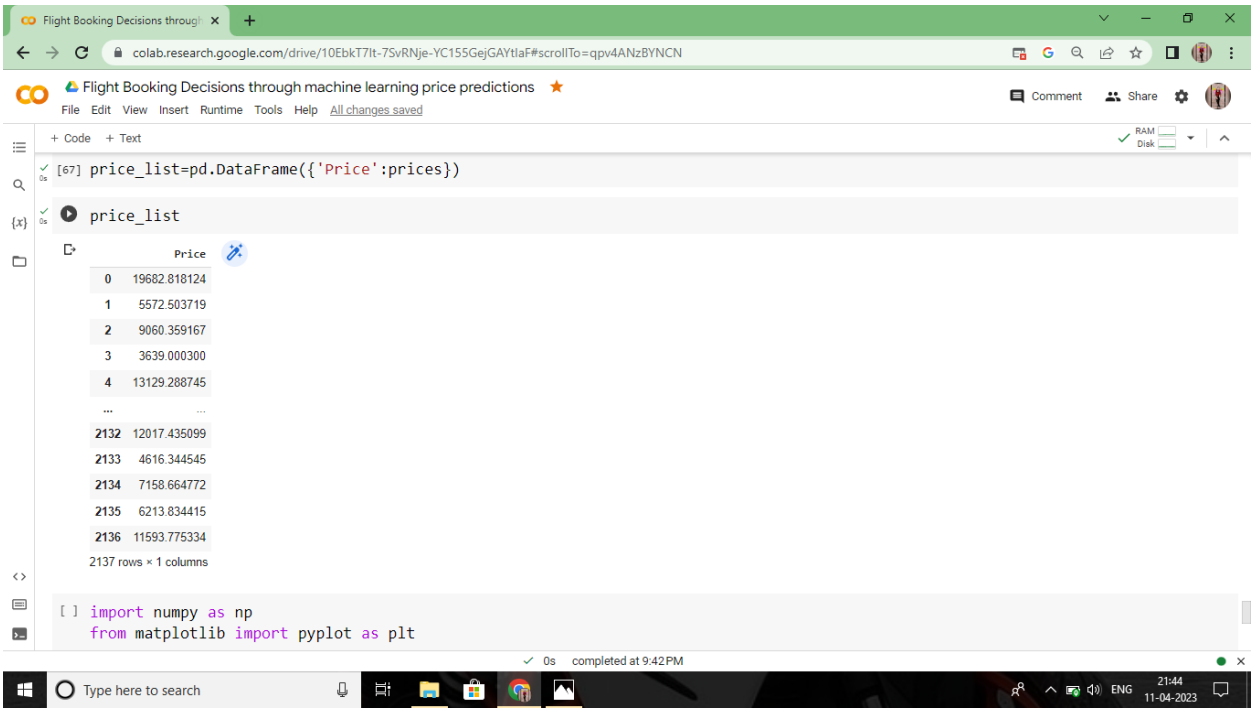

Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
R2 Score is -0.03251945438190873
[74] R2 Score for train data -0.02447186805496515
Mean Squared Error is 3627.1188577608436
Mean Squared Error is 21826451.393833652
Root Mean Squared Error is 4671.878786295044
DecisionTreeRegressor()
R2 Score is 0.6339902382534257
R2 Score for train data 0.7321334660000298
Mean Squared Error is 1746.012685649692
Mean Squared Error is 7737088.381751089
Root Mean Squared Error is 2781.5622196440418
[55] from sklearn.model_selection import cross_val_score
[56] for i in range(2,5):
cv=cross_val_score(rfr,x,y,cv=i)
print(rfr,cv.mean())
RandomForestRegressor() 0.6074847564173703
RandomForestRegressor() 0.6213800672092057
RandomForestRegressor() 0.6380692025000014
[57] from sklearn.model_selection import RandomizedSearchCV
[58] param_grid={'n_estimators':[10,30,50,70,100],'max_depth':[None,1,2,3],'max_features':['auto','sqrt']}
[59] rfr=RandomForestRegressor()
0s completed at 9:42PM
Type here to search
ENG 21:44 11-04-2023


Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
[59] rfr=RandomForestRegressor()
rfr.fit(x_train,y_train)
RandomForestRegressor
RandomForestRegressor()
[60] gb=GradientBoostingRegressor()
gb_res=RandomizedSearchCV(estimator=gb,param_distributions=param_grid,cv=3,verbose=2,n_jobs=-1)
[61] gb_res.fit(x_train,y_train)
Fitting 3 folds for each of 10 candidates, totalling 30 fits
RandomizedSearchCV
estimator: GradientBoostingRegressor
GradientBoostingRegressor
Accuracy
0s completed at 9:42PM
Type here to search
ENG 21:44 11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
gb_res.fit(x_train,y_train)
Fitting 3 folds for each of 10 candidates, totalling 30 fits
RandomizedSearchCV
estimator: GradientBoostingRegressor
GradientBoostingRegressor
Accuracy
[62] rfr=RandomForestRegressor(n_estimators=10,max_features='sqrt',max_depth=None)
rfr.fit(x_train,y_train)
y_train_pred=rfr.predict(x_train)
y_test_pred=rfr.predict(x_test)
print("train accuracy",r2_score(y_train_pred,y_train))
print("test accuracy",r2_score(y_test_pred,y_test))
train accuracy 0.6064933937571388
test accuracy 0.45852312577640075
[75] knn=KNeighborsRegressor(n_neighbors=2,algorithm='auto',metric_params=None,n_jobs=-1)
knn.fit(x_train,y_train)
y_train_pred=knn.predict(x_train)
0s completed at 9:42PM
Type here to search
ENG
21:44
11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
knn=KNeighborsRegressor(n_neighbors=2,algorithm='auto',metric_params=None,n_jobs=-1)
knn.fit(x_train,y_train)
y_train_pred=knn.predict(x_train)
y_test_pred=knn.predict(x_test)
print("train accuracy",r2_score(y_train_pred,y_train))
print("test accuracy",r2_score(y_test_pred,y_test))
train accuracy 0.5082560160438218
test accuracy 0.28385693588182304
[76] predicted_values=pd.DataFrame({'Actual':y_test,'Predicted':y_pred})
[77] predicted_values
Actual Predicted
6075 16655 19069.857143
3544 4959 5496.333333
9291 9187 8928.000000
5032 3858 3657.230769
2483 12898 12821.529412
... ...
9797 7408 12319.470588
0s completed at 9:42PM
Type here to search
ENG
21:44
11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
[77] predicted_values
Actual Predicted
6075 16655 19069.857143
3544 4959 5496.333333
9291 9187 8928.000000
5032 3858 3657.230769
2483 12898 12821.529412
... ...
9797 7408 12319.470588
9871 4622 4903.714286
10063 7452 7104.100000
8803 7060 6244.185185
8618 13731 11612.809524
2137 rows × 2 columns
[66] prices=rfr.predict(x_test)
[67] price_list=pd.DataFrame({'Price':prices})
0s completed at 9:42PM
Type here to search
ENG
21:44
11-04-2023


Flight Booking Decisions through
colab.research.google.com/drive/10EbkT7It-7SvRNje-YC155GejGAYtlaF#scrollTo=qpv4ANzBYNCN
Flight Booking Decisions through machine learning price predictions
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
[67] price_list=pd.DataFrame({'Price':prices})
price_list
Price
0 19682.818124
1 5572.503719
2 9060.359167
3 3639.000300
4 13129.288745
... ...
2132 12017.435099
2133 4616.344545
2134 7158.664772
2135 6213.834415
2136 11593.775334
2137 rows × 1 columns
[ ] import numpy as np
from matplotlib import pyplot as plt
0s completed at 9:42PM
Type here to search
ENG
21:44
11-04-2023

Flight Booking Decisions through machine learning price predictions

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

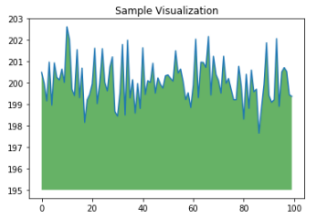
+ Code + Text

```
import numpy as np
from matplotlib import pyplot as plt

ys = 200 + np.random.randn(100)
x = [x for x in range(len(ys))]

plt.plot(x, ys, '-')
plt.fill_between(x, ys, 195, where=(ys > 195), facecolor='g', alpha=0.6)

plt.title("Sample Visualization")
plt.show()
```



0s completed at 9:42PM

Type here to search

ENG 21:44 11-04-2023