1. **INTRODUCTION**

**1.1 Overview**

Those who frequently fly for work will be more knowledgeable about the greatest deals and the best times to purchase tickets. Many airline firms adjust their fares in accordance with the seasons or time periods for commercial reasons. When more people go, the cost will go up. Data for the route is gathered using features like Duration, Source, Destination, Arrival, and Departure to estimate the highest airline fares. Features are gathered from a chosen dataset and are included in the cost, where the price of an airline ticket changes over time. We used KNN, decision trees, and random forests to implement flight price prediction for users. For estimating the cost of a flight, Random Forest has the highest accuracy (80%). Moreover, we conducted correlation analyses and metrics. Nowadays, the majority of ancillary pricing decisions are made manually by analysts using historical data analysis and benchmarking against competitors. Merchandising rules allow ancillary pricing to be further customised to the specifics of the ancillary request after they have been manually calculated and entered in ATPCO (Airline Tariff Publishing Company) or Merchandising systems. We developed a random forest machine algorithm using airline ancillary and itinerary data that can comprehend the complex relationships between numerous attributes, such as passenger type, itinerary, aircraft type, ancillary product, or season, and can automatically determine pricing based on science.

**Key words: Random Forest, flight price prediction, KNN, decision trees.**

**1.2 Purpose**

**Bisiness problem:**

Flight price prediction is an essential tool that airlines use to enhance customer satisfaction and loyalty. By analyzing vast amounts of data on customer behavior, past purchase history, and other relevant information, airlines can develop personalized pricing and offers that cater to the unique needs of individual customers. One of the primary benefits of flight price prediction is that it enables airlines to optimize their revenue streams. By offering personalized pricing and offers, airlines can maximize the value of each customer transaction and ensure that customers are willing to pay the best price for their travel needs. This can help airlines to increase revenue, improve profitability, and compete more effectively in the marketplace.

Another benefit of flight price prediction is that it helps to improve customer satisfaction. By tailoring pricing and offers to individual customers, airlines can provide a more personalized and convenient experience that meets their specific needs. This can include offering discounts or rewards to frequent flyers or providing customized travel packages that include preferred seats, meals, and other amenities. Moreover, flight price prediction can also help airlines to build stronger customer relationships. By leveraging customer data, airlines can identify the unique preferences and travel patterns of their customers and use this information to develop targeted marketing campaigns and promotions. This can help airlines to engage with customers more effectively and build long-term loyalty.

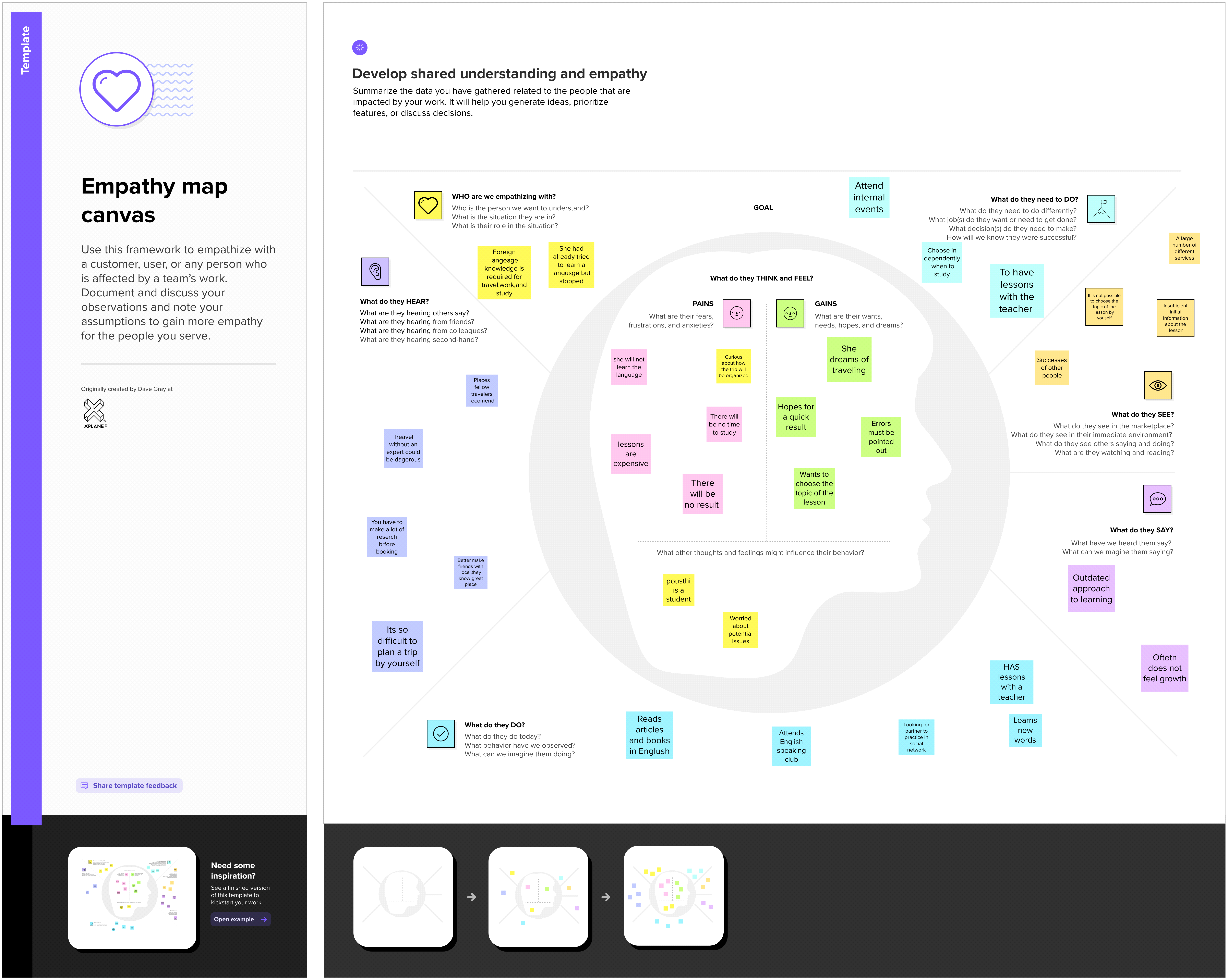
In summary, flight price prediction is a powerful tool that can help airlines to enhance customer satisfaction and drive revenue growth. By leveraging customer data and advanced analytical techniques, airlines can develop personalized pricing and offers that cater to the unique needs of individual customers, build stronger relationships, and compete more effectively in the marketplace.

**Business requirement:**

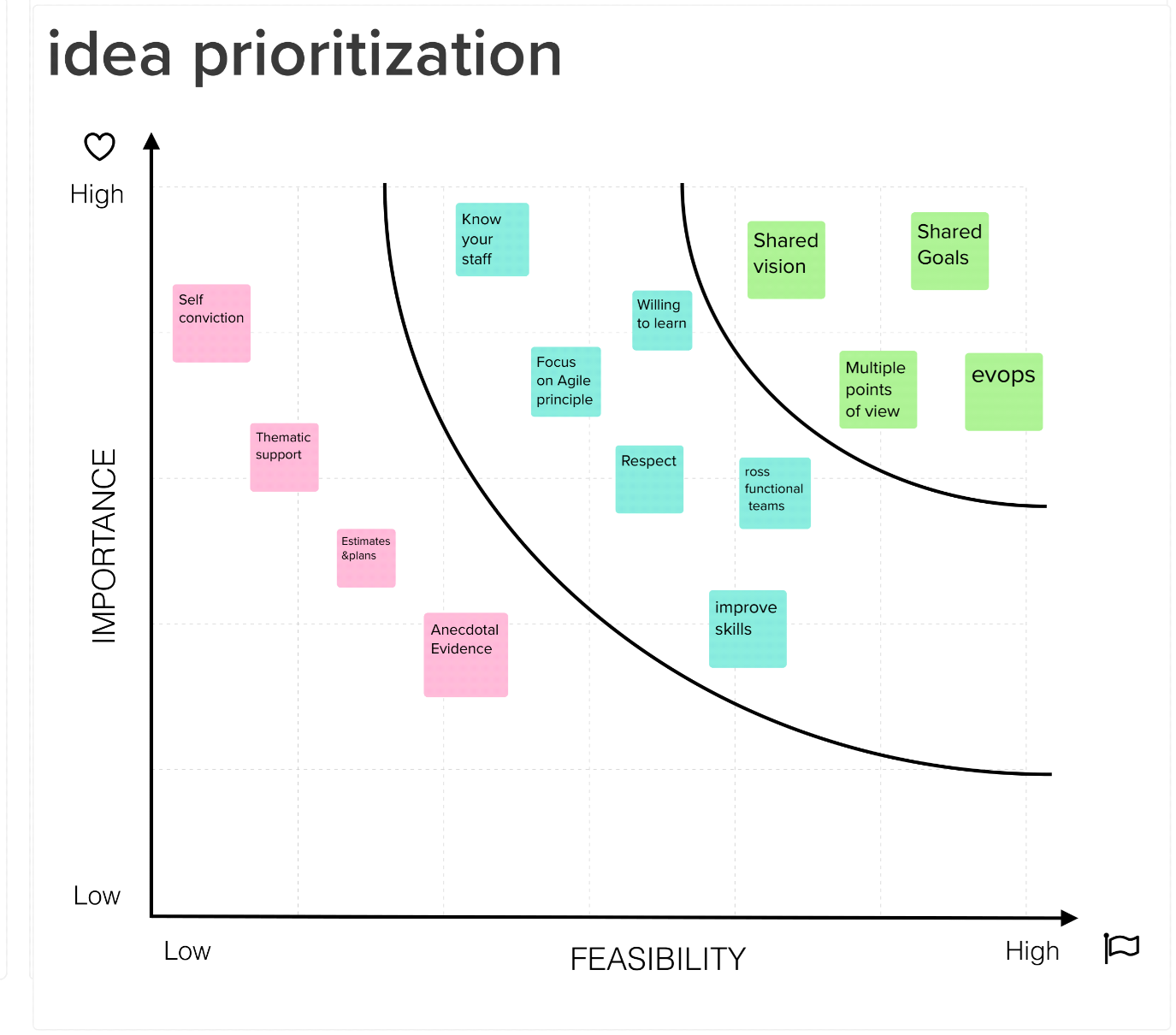
1. To develop a flight price prediction system that can analyze customer data, past purchase history, loyalty program status, and demographic information to develop personalized pricing and offers for individual customers. The system should be able to optimize revenue streams, improve customer satisfaction, and build stronger customer relationships by offering tailored pricing and promotions that meet the unique needs of individual customers. Additionally, the system should be able to provide airlines with insights into customer behavior, preferences, and travel patterns, which can be used to develop targeted marketing campaigns and promotions. The system should be reliable, accurate, and scalable to handle large volumes of data and support the needs of a global airline network.

**2.Problem Definition & Design Thinking**

**2.1 Empathy Map**



**2.2 Ideation & Brinstorming map**



1. **RESULT**

**Result:1**

Airline ['IndiGo' 'Air India' 'Jet Airways' 'SpiceJet' 'Multiple carriers' 'GoAir'

'Vistara' 'Air Asia' 'Vistara Premium economy' 'Jet Airways Business'

'Multiple carriers Premium economy']

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']

**Result:2**

array(['No info', 'In-flight meal not included',

'No check-in baggage included', '1 Short layover', 'No Info',

'1 Long layover'], dtype=object)

**Result:3**

<bound method NDFrame.\_add\_numeric\_operations.<locals>.sum of Airline Date\_of\_Journey Source Destination Route Dep\_Time \

0 False False False False False False

1 False False False False False False

2 False False False False False False

3 False False False False False False

4 False False False False False False

... ... ... ... ... ... ...

1285 False False False False False False

1286 False False False False False False

1287 False False False False False False

1288 False False False False False False

1289 False False False False False False

Arrival\_Time Duration Total\_Stops Additional\_Info ... Month Year \

0 False False False False ... True True

1 False False False False ... True True

2 False False False False ... True True

3 False False False False ... True True

4 False False False False ... True True

... ... ... ... ... ... ... ...

1285 False False False False ... True True

1286 False False False False ... True True

1287 False False False False ... True True

1288 False False False False ... True True

1289 False False False False ... True True

Dep\_Time\_Hour Dep\_Time\_Mins Arrival\_date Time\_of\_Arrival \

0 False False False False

1 False False True False

2 False False False False

3 False False True False

4 False False True False

... ... ... ... ...

1285 False False True False

1286 False False True False

1287 False False True False

1288 False False True False

1289 False False True False

Arrival\_Time\_Hour Arrival\_Time\_Mins Travel\_Hours Travel\_Mins

0 False True False False

1 False False False False

2 False True False False

3 False False False False

4 False False False False

... ... ... ... ...

1285 False False False False

1286 False False False False

1287 False False False False

1288 False False False False

1289 False False False False

[1290 rows x 28 columns]>

**Result:4**

Index(['Airline', 'Date\_of\_Journey', 'Source', 'Destination', 'Route',

'Dep\_Time', 'Arrival\_Time', 'Duration', 'Total\_Stops',

'Additional\_Info', 'Price', 'City1', 'City2', 'City3', 'Date', 'Month',

'Year', 'Dep\_Time\_Hour', 'Dep\_Time\_Mins', 'Arrival\_date',

'Time\_of\_Arrival', 'Arrival\_Time\_Hour', 'Arrival\_Time\_Mins',

'Travel\_Hours', 'Travel\_Mins'],

dtype='object')

<class 'pandas.core.frame.DataFrame'>

Int64Index: 1290 entries, 0 to 1289

Data columns (total 19 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Airline 1290 non-null object

1 Source 1290 non-null object

2 Destination 1290 non-null object

3 Total\_Stops 1290 non-null object

4 Additional\_Info 1290 non-null object

5 Price 1290 non-null float64

6 City1 1290 non-null object

7 City2 0 non-null float64

8 City3 0 non-null float64

9 Date 1290 non-null object

10 Month 0 non-null float64

11 Year 0 non-null float64

12 Dep\_Time\_Hour 1290 non-null object

13 Dep\_Time\_Mins 1290 non-null object

14 Arrival\_date 536 non-null object

15 Arrival\_Time\_Hour 1290 non-null object

16 Arrival\_Time\_Mins 754 non-null object

17 Travel\_Hours 1290 non-null object

18 Travel\_Mins 1290 non-null object

dtypes: float64(5), object(14)

memory usage: 201.6+ KB

Airline Date\_of\_Journey Source Destination Route Dep\_Time Arrival\_Time Duration Total\_Stops Additional\_Info ... Month Year Dep\_Time\_Hour Dep\_Time\_Mins Arrival\_date Time\_of\_Arrival Arrival\_Time\_Hour Arrival\_Time\_Mins Travel\_Hours Travel\_Mins

0 rows × 28 columns

**Result:5**

**KeyError: 'Additional Info'**

**The above exception was the direct cause of the following exception:**

**KeyError Traceback (most recent call last)**

**/usr/local/lib/python3.9/dist-packages/pandas/core/indexes/base.py in get\_loc(self, key, method, tolerance)**

**3629 return self.\_engine.get\_loc(casted\_key)**

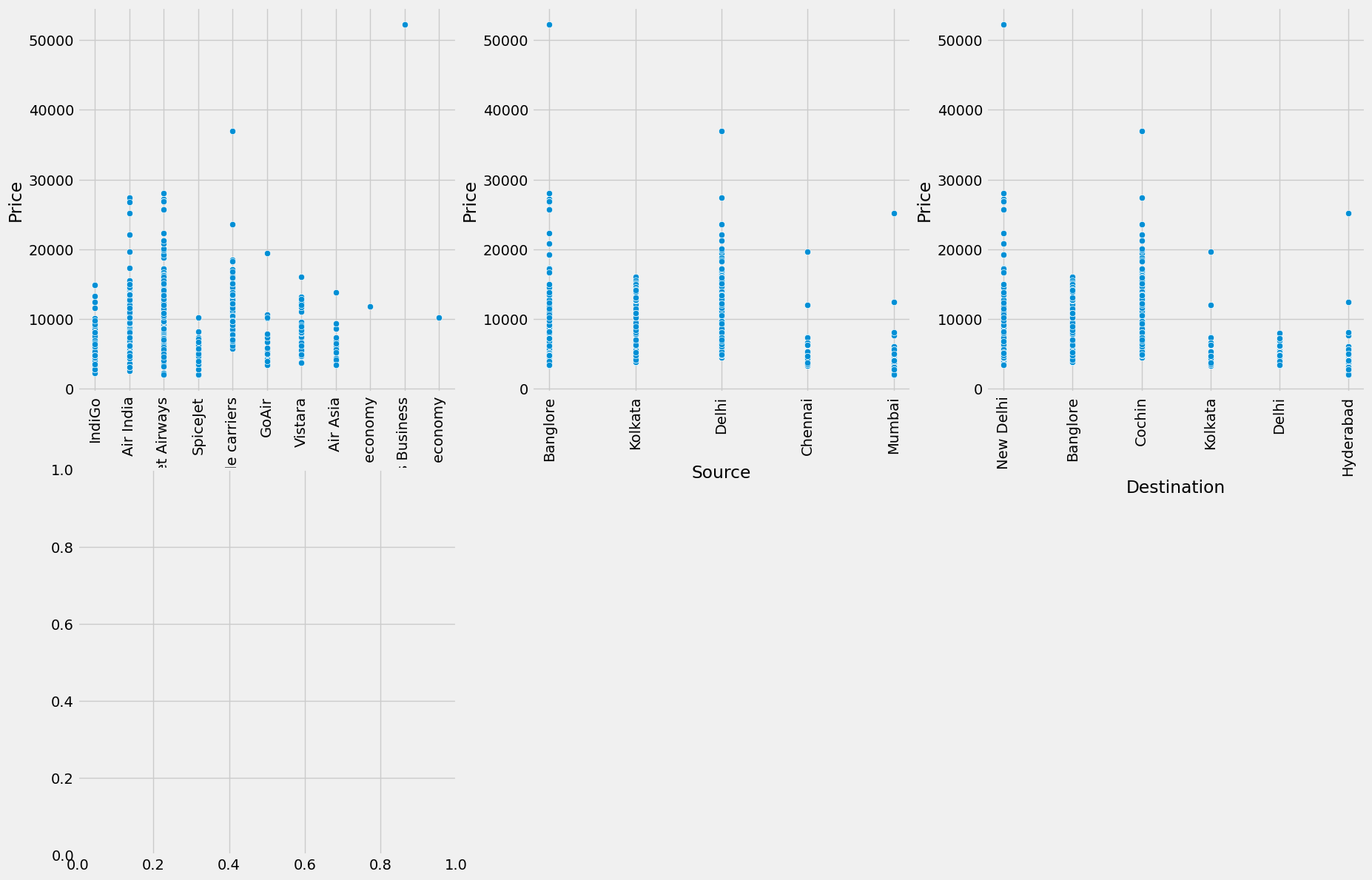
**3630 except KeyError as err:**

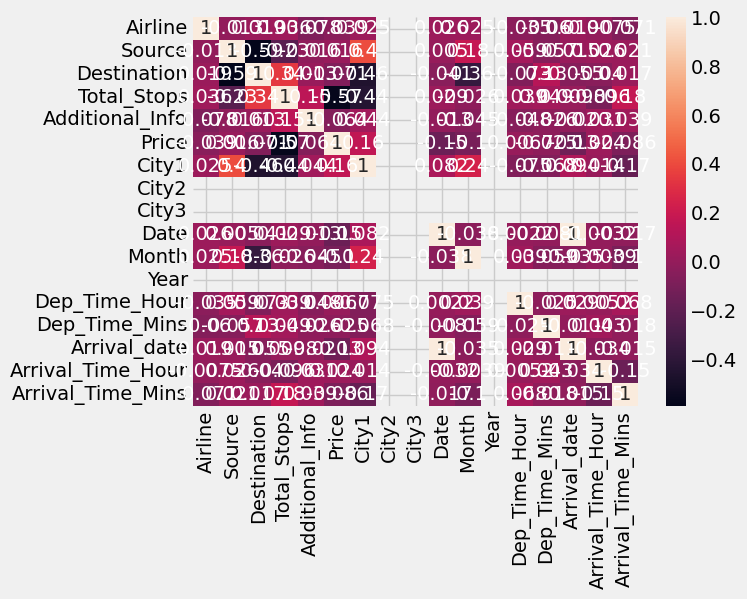
**-> 3631 raise KeyError(key) from err**

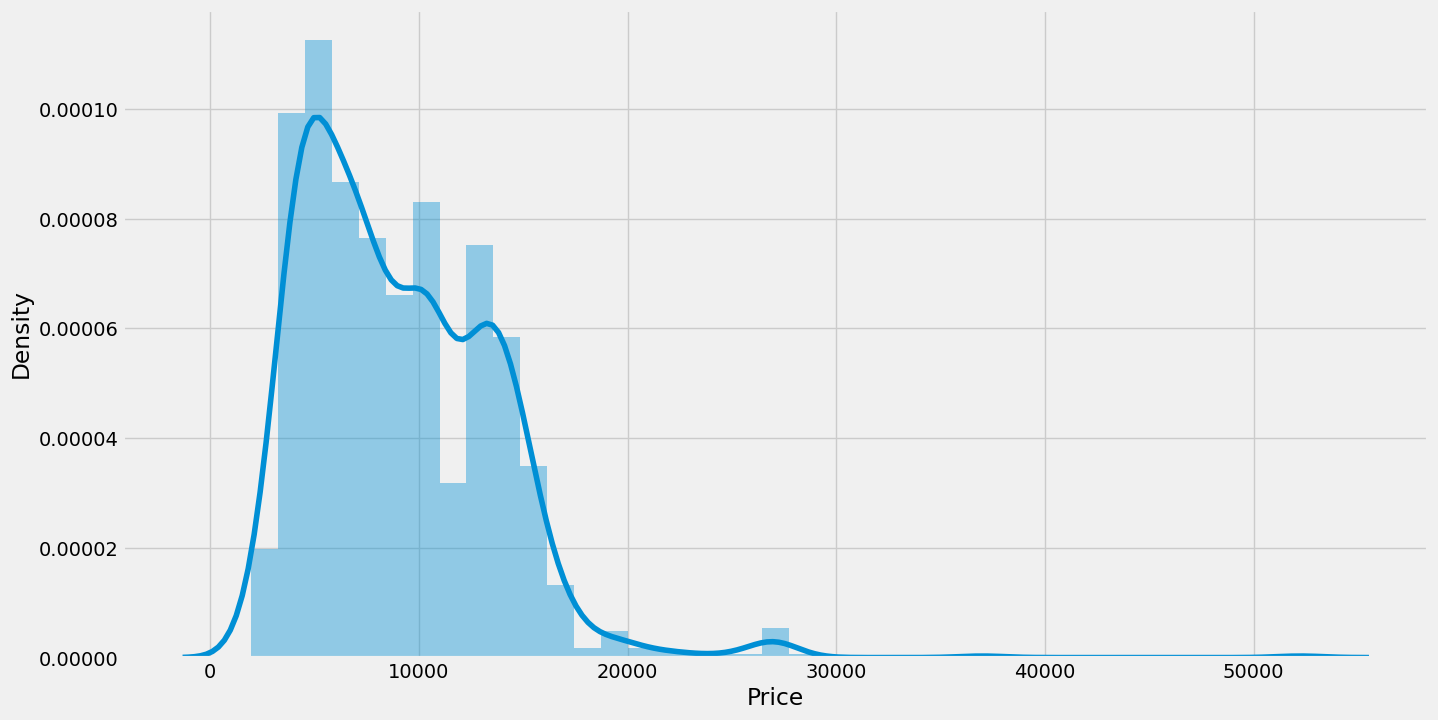
**3632 except TypeError:**

**3633 # If we have a listlike key, \_check\_indexing\_error will raise**

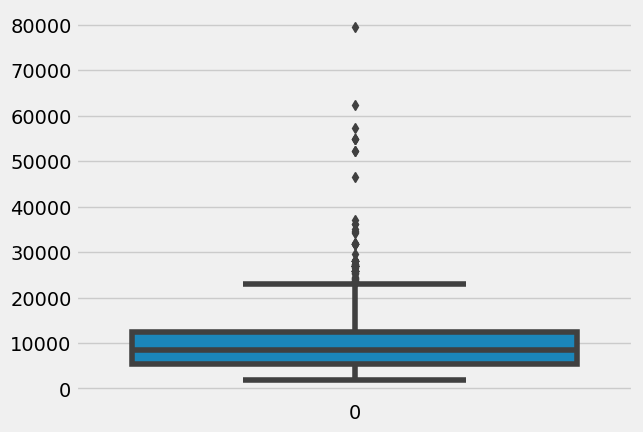
**KeyError: 'Additional Info'**

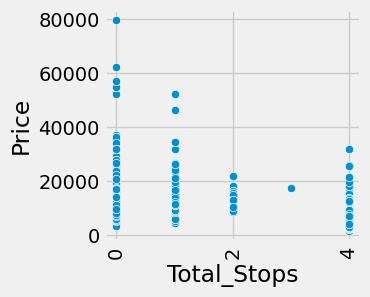


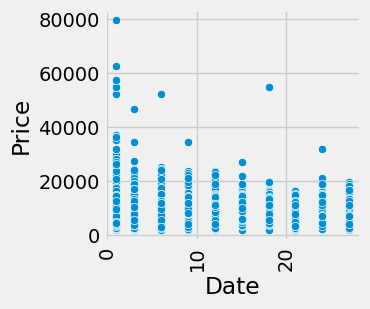


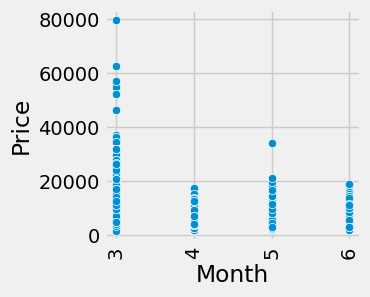


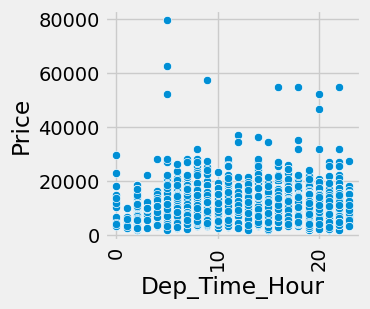
**Result:6**

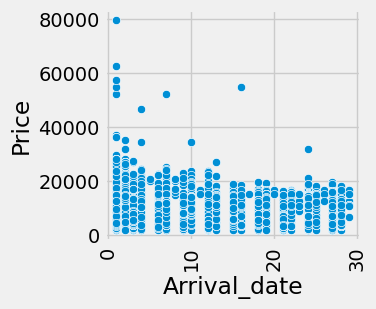


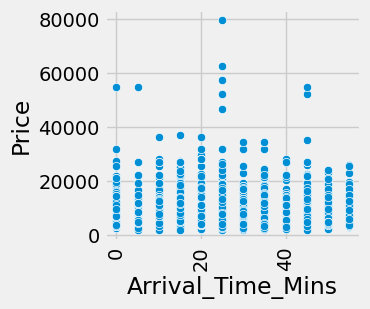












**Result:7**

**Airline Source Destination Date Month Year Dep\_Time\_Hour Dep\_Time\_Mins Arrival\_date Arrival\_Time\_Hour Arrival\_Time\_Mins Price**

**0 3 0 5 24 3 2019 22 20 22 1 10 3897**

**1 1 3 0 1 5 2019 5 50 1 13 15 7662**

**2 4 2 1 9 6 2019 9 25 10 4 25 13882**

**3 3 3 0 12 5 2019 18 5 12 23 30 6218**

**4 3 0 5 1 3 2019 16 50 1 21 35 13302**

**index Airline Source Destination Date Month Year Dep\_Time\_Hour Dep\_Time\_Mins Arrival\_date Arrival\_Time\_Hour Arrival\_Time\_Mins Price**

**0 3 0 5 24-03-2019 NaN NaN 22 20 01:10 22-03-2023 NaN 3897.0**

**1 1 3 0 01-05-2019 NaN NaN 05 50 01-05-2019 13 15 7662.0**

**2 4 2 1 09-06-2019 NaN NaN 09 25 04:25 10-06-2023 NaN 13882.0**

**3 3 3 0 12-05-2019 NaN NaN 18 05 12-05-2019 23 30 6218.0**

**4 3 0 5 01-03-2019 NaN NaN 16 50 01-03-2019 21 35 13302.0**

**Show**

**25**

**per page**

**Like what you see? Visit the data table notebook to learn more about interactive tables.**

**Airline Source Destination Date Month Year Dep\_Time\_Hour Dep\_Time\_Mins Arrival\_date Arrival\_Time\_Hour Arrival\_Time\_Mins Price**

**0 -0.410934 -1.658354 2.416648 1.237192 -1.467619 0.0 1.654162 -0.234832 0.955658 -1.800328 -0.889941 -1.125483**

**1 -1.261305 0.890262 -0.973718 -1.475375 0.250165 0.0 -1.303018 1.363790 -1.524701 -0.050871 -0.586988 -0.308932**

**2 0.014251 0.040723 -0.295645 -0.531874 1.109057 0.0 -0.607211 0.031605 -0.461690 -1.362964 0.018919 1.040057**

**3 -0.410934 0.890262 -0.973718 -0.178060 0.250165 0.0 0.958355 -1.034142 -0.225465 1.407010 0.321872 -0.622106**

**4 -0.410934 -1.658354 2.416648 -1.475375 -1.467619 0.0 0.610452 1.363790 -1.524701 1.115434 0.624825 0.914267**

**Result:8Model: "sequential"**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Layer (type) Output Shape Param #**

**=================================================================**

**dense (Dense) (None, 7) 84**

**dense\_1 (Dense) (None, 7) 56**

**dense\_2 (Dense) (None, 1) 8**

**=================================================================**

**Total params: 148**

**Trainable params: 148**

**Non-trainable params: 0**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**RandomForestRegressor() 0.7909852737870107**

**RandomForestRegressor() 0.7935431966039664**

**RandomForestRegressor() 0.8004797889513**

**Fitting 3 folds for each of 10 candidates, totalling 30 fits**

**RandomizedSearchCV**

**estimator: RandomForestRegressor**

**Result:9**

**RandomForestRegressorRandomForestRegressor(max\_features='sqrt', n\_estimators=10) 0.72661809392105**

**RandomForestRegressor(max\_features='sqrt', n\_estimators=10) 0.7287548229046766**

**RandomForestRegressor(max\_features='sqrt', n\_estimators=10) 0.728029951483208**

**KNeighborsRegressor()**

**R2 Score is 0.7357369816529409**

**R2 Score for train data 0.7900498333828809**

**Mean Absolute Error is 0.35463454315938664**

**Mean Squared Error is 0.26242008660326566**

**Root Mean Squared Error is 0.512269544871902**

**SVR()**

**R2 Score is 0.6399736388140904**

**R2 Score for train data 0.5969176412610055**

**Mean Absolute Error is 0.40820604052912457**

**Mean Squared Error is 0.3575155898574727**

**Root Mean Squared Error is 0.5979260739066935**

**KNeighborsRegressor(n\_jobs=-1, n\_neighbors=2) 0.6306338018391912**

**KNeighborsRegressor(n\_jobs=-1, n\_neighbors=2) 0.6447308601134175**

**KNeighborsRegressor(n\_jobs=-1, n\_neighbors=2) 0.664555765507016**

**Actual Predicted**

**4830 -0.349272 -0.455760**

**3771 -0.251459 -0.171648**

**1523 -0.677410 0.638830**

**3393 1.562086 0.826648**

**4169 -0.232157 -0.719485**

**... ... ...**

**9869 -0.968245 -0.614949**

**10061 -0.354477 -0.354477**

**6911 -0.348404 -0.348404**

**8616 1.098398 1.138303**

**8988 1.254551 1.254551**

**Price**

**0 -0.541116**

**1 -0.057338**

**2 0.496086**

**3 0.964626**

**4 -0.683179**

**... ...**

**2132 -0.614212**

**2133 -0.549343**

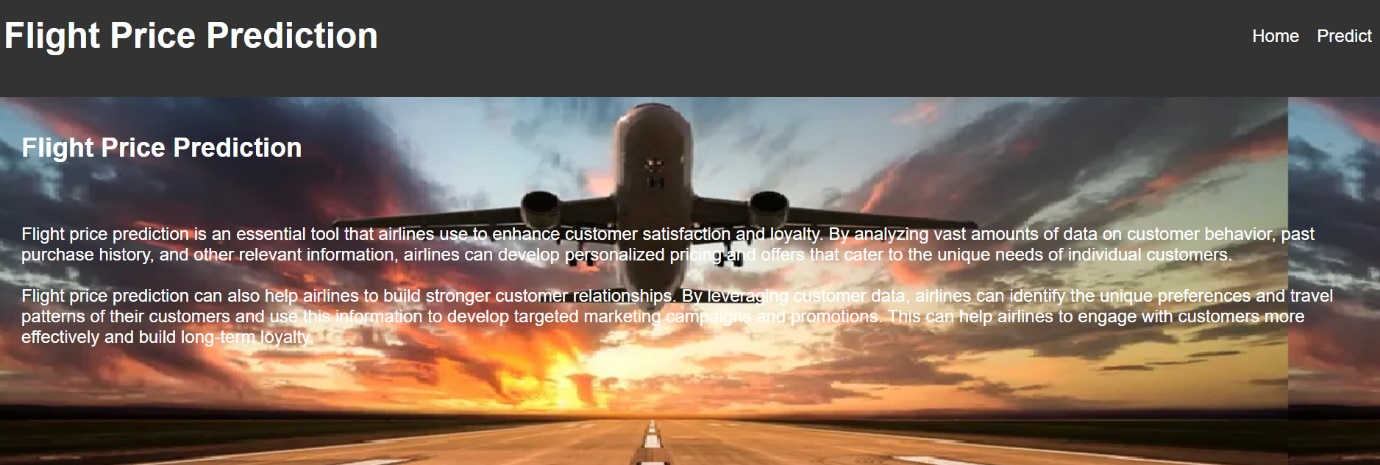
**2134 -0.374603**

**2135 0.738894**

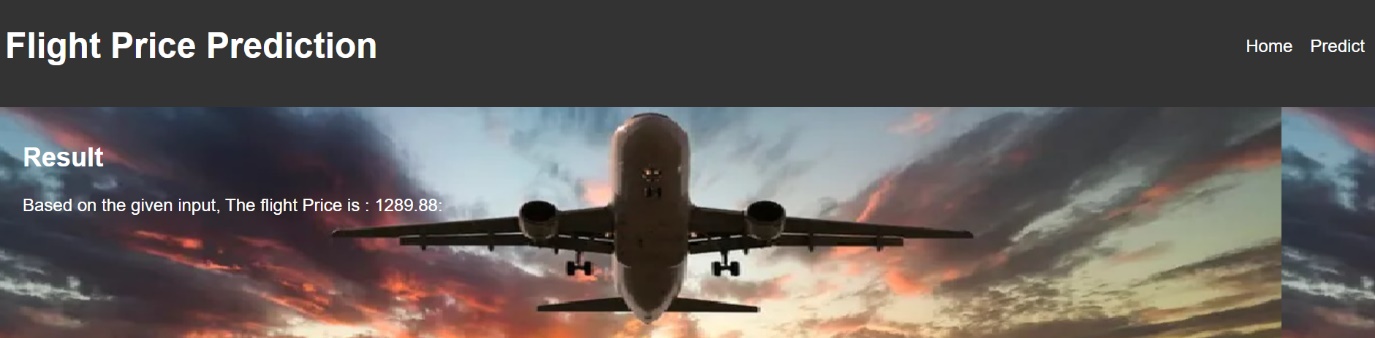
**2136 1.056930**

**2137 rows × 1 columns**

**Output Screen**

****

****

****

1. **ADVANTAGE AND DISADVANTAGE**

**Advantages**

Increased customer satisfaction: Personalized pricing and offers can enhance the travel experience by providing customers with tailored options that meet their individual needs and preferences.

Improved revenue: Optimized pricing strategies can increase revenue and profitability by maximizing the value of each customer transaction.

Enhanced loyalty: Personalized pricing and offers can help build stronger customer relationships by showing customers that the airline values their business and is willing to provide customized services.

Better marketing: Analyzing customer data can provide airlines with insights into customer behavior and preferences, allowing them to develop targeted marketing campaigns and promotions that engage customers more effectively.

Competitive advantage: By providing personalized pricing and offers, airlines can differentiate themselves from their competitors and improve their market position.

**Disadvantages** of using flight price prediction methods to personalize pricing and offers include:

Privacy concerns: Customers may be hesitant to share personal information with airlines, particularly if they feel that their data is not being used appropriately or ethically.

Accuracy: Predictive models may not always be accurate, and there is a risk of mispricing or misidentifying customer preferences.

Complexity: Developing and implementing a flight price prediction system can be complex and requires significant resources, including expertise in data analytics and machine learning.

Regulatory compliance: The use of customer data is subject to regulatory compliance requirements, which can be complex and time-consuming to navigate.

Customer dissatisfaction: Personalized pricing and offers may not always be perceived as fair or transparent, leading to customer dissatisfaction and potential reputational damage for the airline.

1. **APPLICATION**

The application of flight price prediction methods to personalize pricing and offers is widespread within the airline industry. Airlines use these methods to optimize their revenue streams, improve customer satisfaction, and build stronger customer relationships. Here are some specific applications:

Dynamic pricing: Airlines use flight price prediction models to dynamically adjust ticket prices based on real-time demand, availability, and other factors. This can help airlines to optimize revenue and fill more seats.

Personalized offers: Airlines use customer data to create personalized offers, such as discounts, upgrades, and loyalty rewards, based on the customer's past purchase history and preferences.

Ancillary revenue: Airlines use flight price prediction models to analyze customer data and identify opportunities to generate additional revenue from ancillary services, such as baggage fees, seat upgrades, and in-flight services.

Customer segmentation: Airlines use flight price prediction models to segment customers based on their travel behavior and preferences. This can help airlines to develop targeted marketing campaigns and promotions that resonate with specific customer segments.

Competitive analysis: Airlines use flight price prediction models to analyze their competitors' pricing strategies and adjust their own pricing accordingly. This can help airlines to stay competitive in the marketplace and attract customers away from their rivals.

Overall, the application of flight price prediction methods has revolutionized the way airlines price and sell their products and services. By leveraging customer data and advanced analytical techniques, airlines can create personalized pricing and offers that meet the unique needs and preferences of individual customers, drive revenue growth, and build stronger customer relationships.

1. **CONCLUSION**

Flight price prediction methods that leverage customer data to personalize pricing and offers have the potential to enhance customer satisfaction, optimize revenue, and build stronger customer relationships within the airline industry. By analyzing customer data and applying advanced analytical techniques, airlines can create personalized pricing and offers that meet the unique needs and preferences of individual customers. This, in turn, can help airlines to improve customer satisfaction, increase revenue, and compete more effectively in the marketplace.

1. **FUTURE SCOPE**

The future scope of flight price prediction methods is vast. With the increasing availability of customer data and advancements in machine learning algorithms, airlines can continue to improve their pricing strategies and develop even more personalized offers for their customers. In the future, flight price prediction models may also incorporate other factors, such as weather conditions, traffic patterns, and event schedules, to create even more accurate pricing models. Additionally, the use of flight price prediction models may expand beyond the airline industry to other travel sectors, such as hotels and rental cars, to create a more seamless and personalized travel experience for customers. Overall, the future of flight price prediction methods is promising, and their continued development and application are likely to drive significant benefits for both airlines and their customers

1. **APPANDIX**

**CODE**