DSCP Project: calculate the flight ticket price

Haorui Wu, Will Wang, Xiangcheng Li, Xinrui Zhong, Yiming Zhang

Group 6

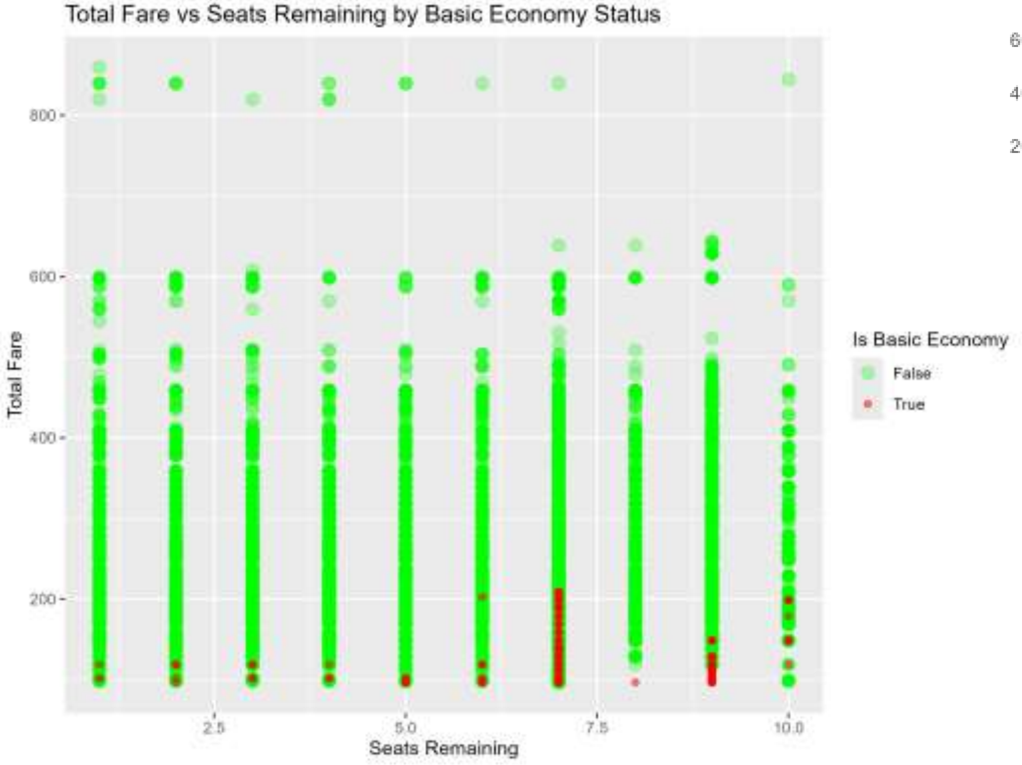
Experiment design

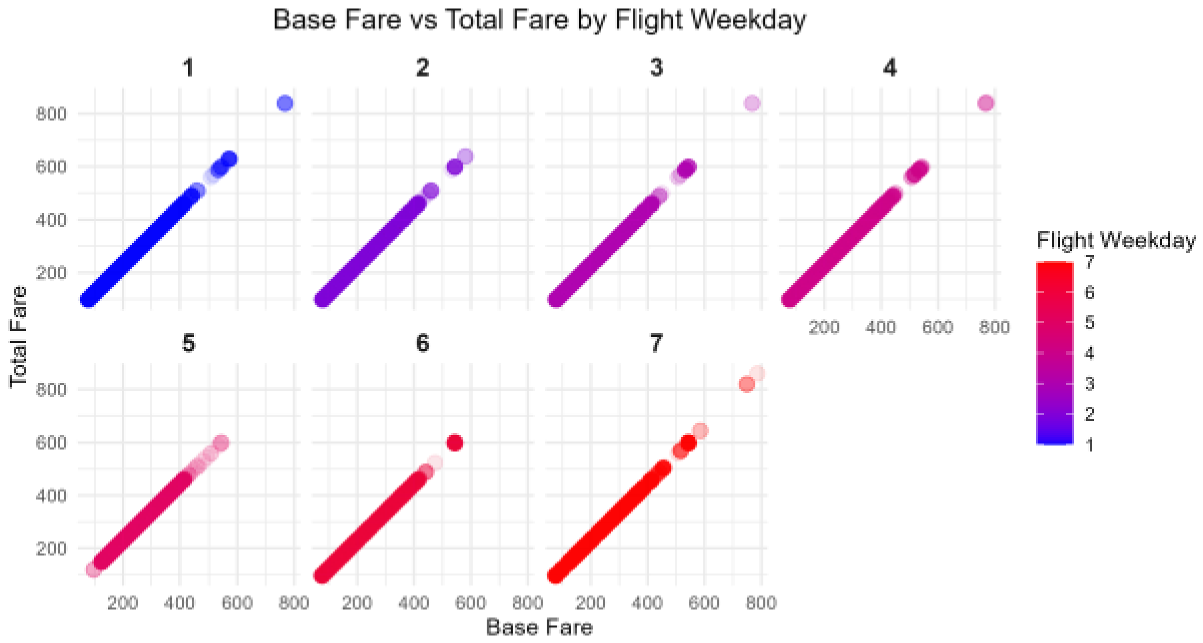
- We get the Flight Price dataset from Kaggle, Flight Prices, to build models to predict the flight price from one airport to another
- Separate into two groups, one data cleaning group and one group focus on parallel computing using CHTC

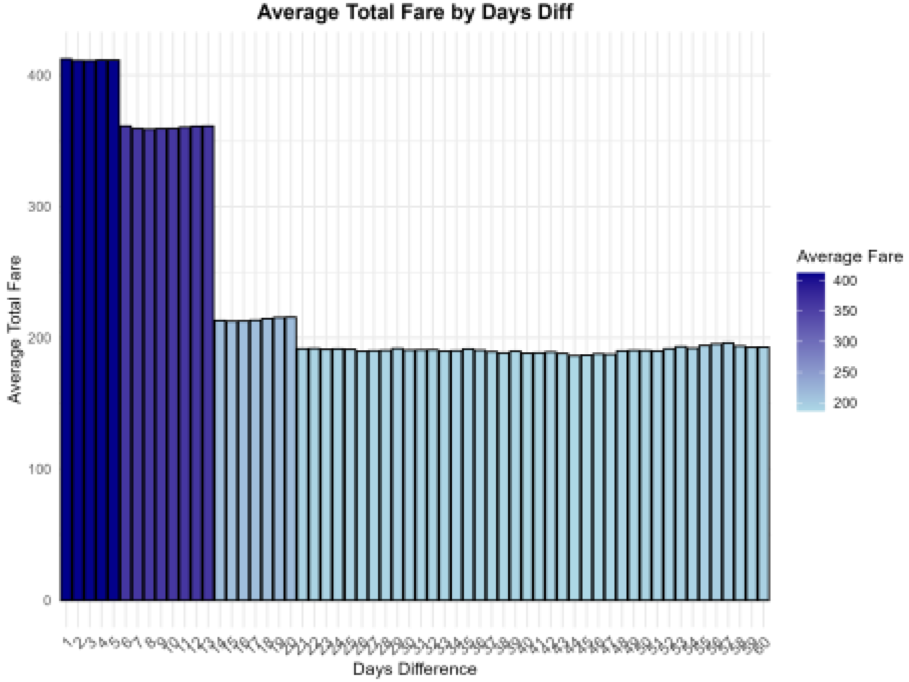
Data Cleaning

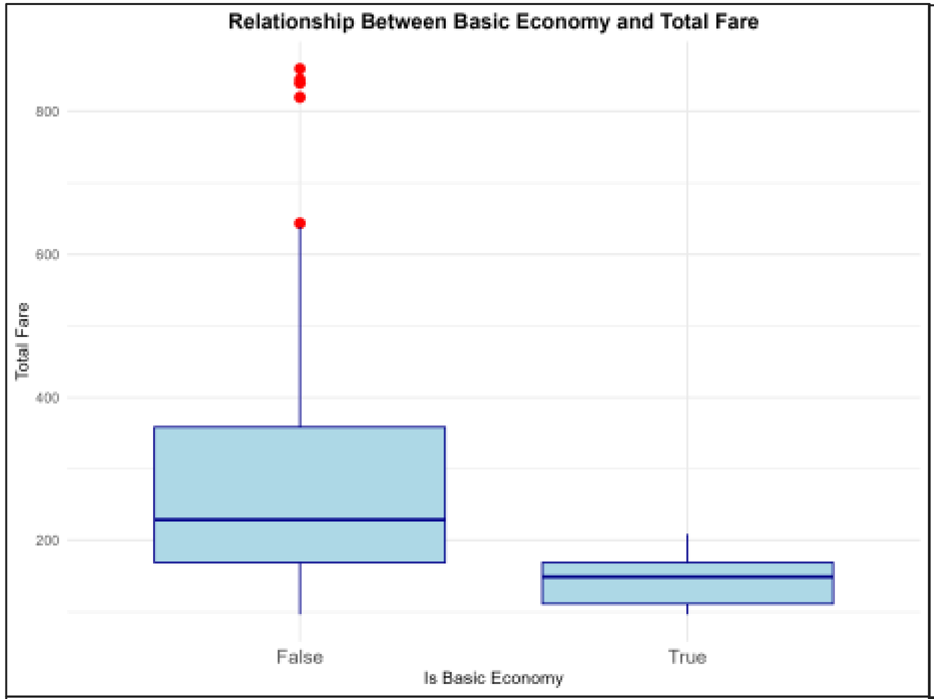
- We downloaded the dataset from Kaggle and remove all the business/first class ticket price since there are few data about the ticket price of the business/first class, we don’t have enough data to support our result if we want to calculate that price.
- We divide the whole big dataset by using the feature startingAirport and destinationAirport which can indicate where the passages’ origins and destinations
- We divided each of the small dataset into two part using the variable isNonStop, this feature indicate if the flight contains a stop or not, which will also make a big influence on the ticket price.

Data Visualization

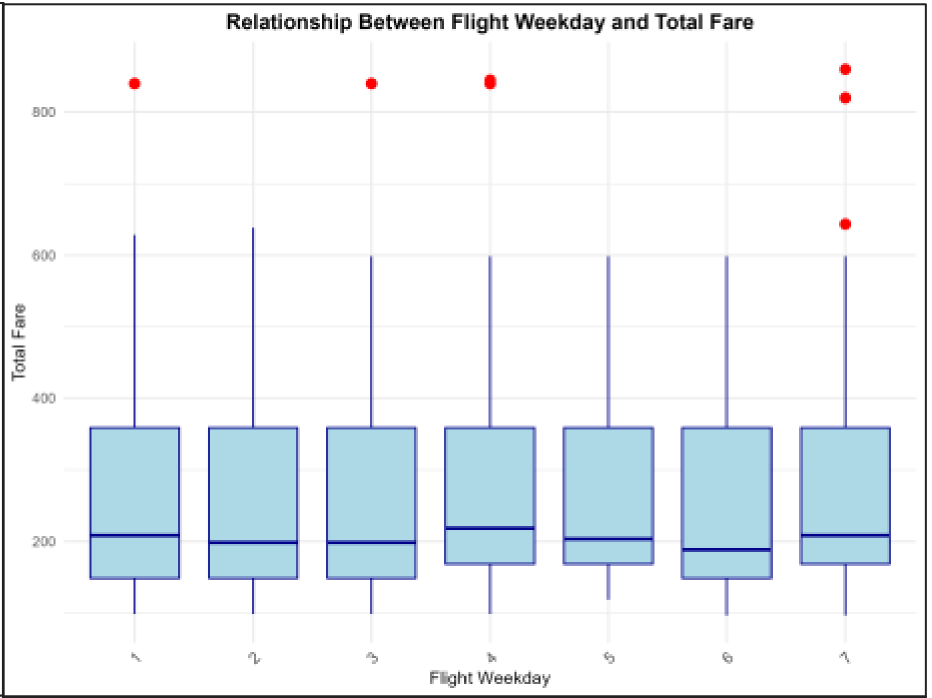

Total Fare vs Seats Remaining by Basic Economy Status


Base Fare vs Total Fare by Flight Weekday

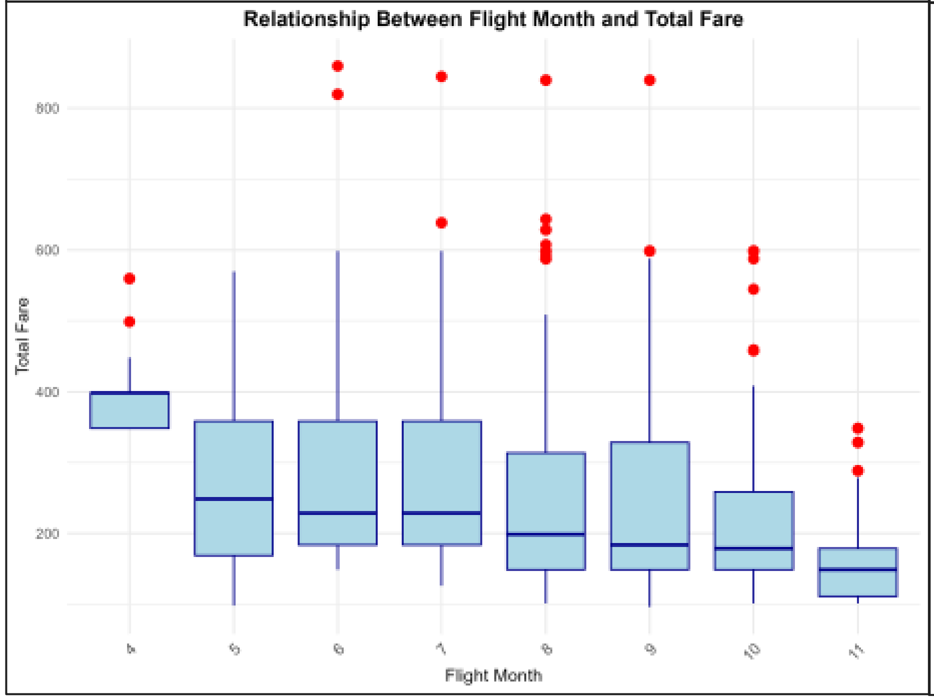

Average Total Fare by Days Diff



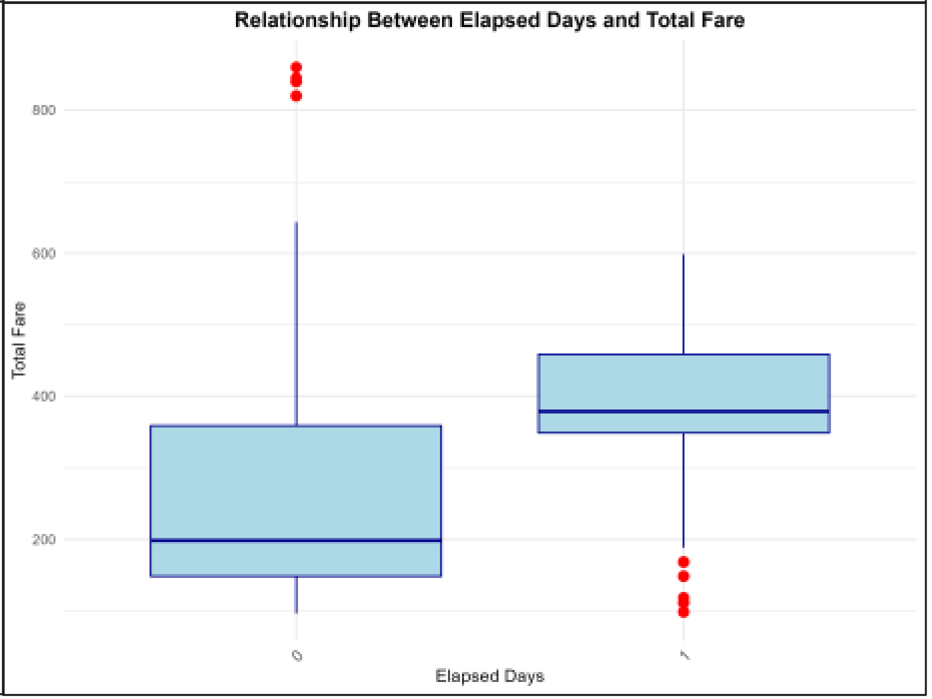
Relationship Between Basic Economy and Total Fare



Relationship Between Flight Weekday and Total Fare

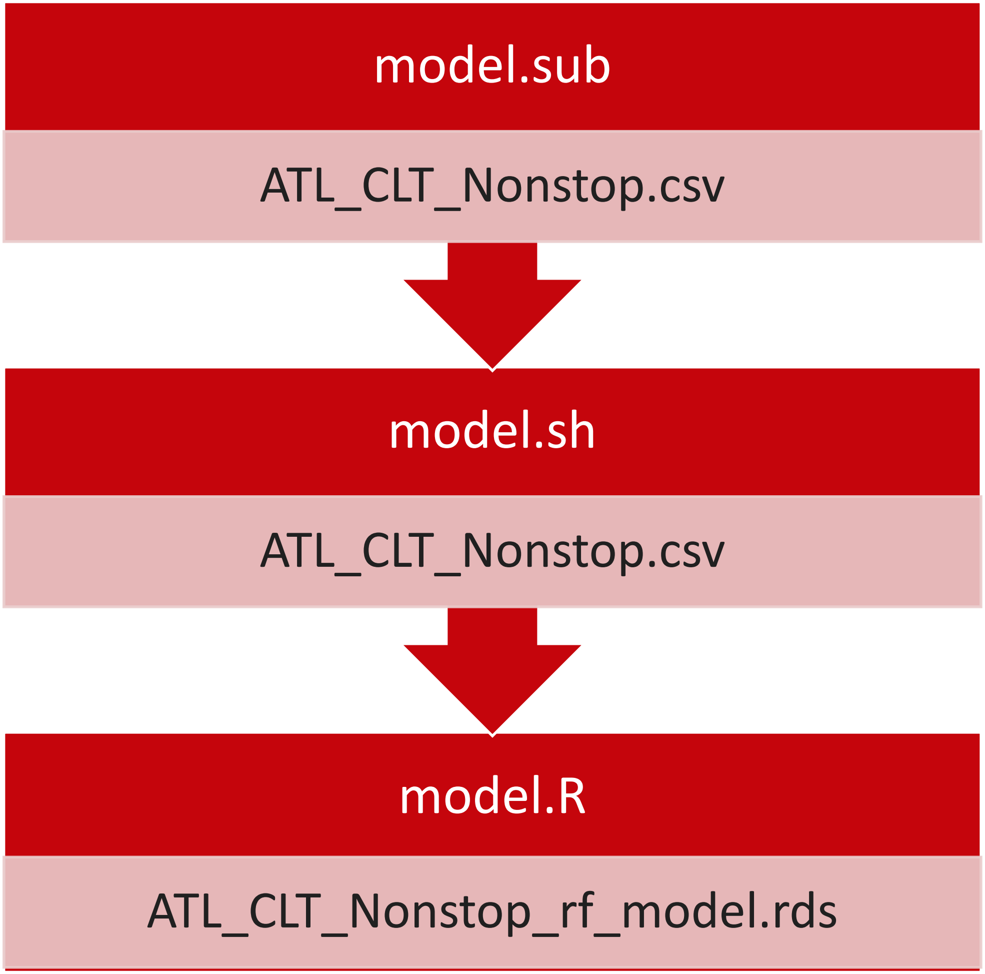


Relationship Between Flight Month and Total Fare



Relationship Between Elapsed Days and Total Fare

Parallel computing



model.sub

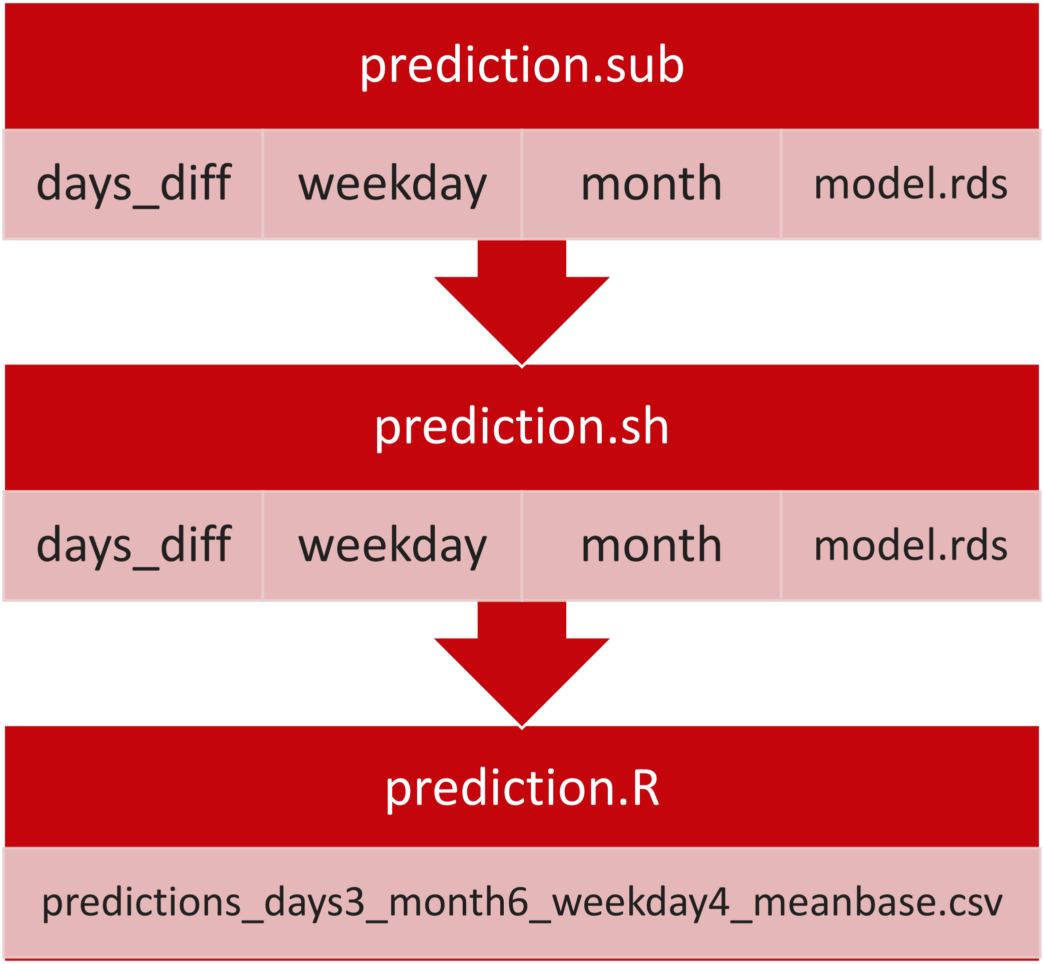
ATL_CLT_Nonstop.csv

model.sh

ATL_CLT_Nonstop.csv

model.R

ATL_CLT_Nonstop_rf_model.rds



prediction.sub

days_diff	weekday	month	model.rds

prediction.sh

days_diff	weekday	month	model.rds

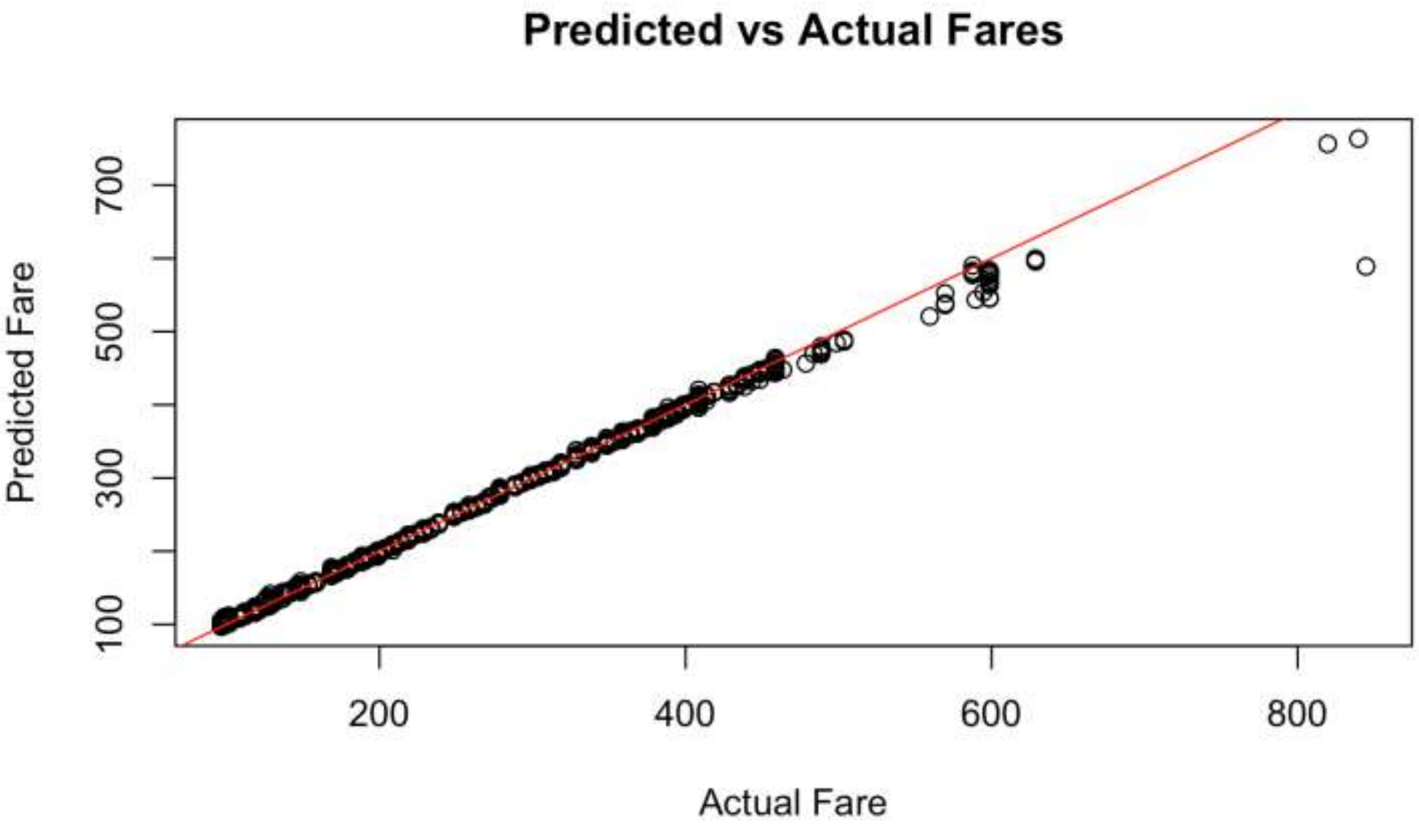
prediction.R

predictions_days3_month6_weekday4_meanbase.csv

Model.R :Random forest model

Feature	Importance
baseFare	119.1
flight_weekday	30.9
seatsRemaining	28.8
flight_month	24.9
isBasicEconomy	18.9
days_diff	17.0
elapsedDays	11.2

Random forest model



Outcome

Base Fare	Elapsed Days	IsBasic Economy	Is Refundable	Seats Remaining	Days Diff	Flight Month	Flight Weekday	Predicted Fare
84.5	0	TRUE	FALSE	9	3	6	4	148.9
76.3	0	TRUE	FALSE	8	3	6	4	149.3
212.2	1	TRUE	FALSE	10	3	6	4	159
191	0	FALSE	FALSE	1	3	6	4	225.5
318.3	1	FALSE	FALSE	4	3	6	4	376.5

WISCONSIN

UNIVERSITY OF WISCONSIN–MADISON