Demand Forecasting

Why demand forecasting ?

- Demand forecasting is a critical component of **revenue** management in any airline

- The Law of Demand states that for most goods, with a few exceptions (Giffen goods, Veblen goods, basic or necessary goods), demand is **inversely** related to price.

- However, the presence of multiple factors often makes the estimation process complicated.

From the demand data

fare	bookings	seasonindex
143.88	3	Off-Peak
167.78	2	Peak
195.21	0	Peak
198.84	4	Peak

.
.
.

- **fare**: continuous data (float)
- **bookings**: continuous data (int)
- **seasonindex**: categorical data (boolean / int)

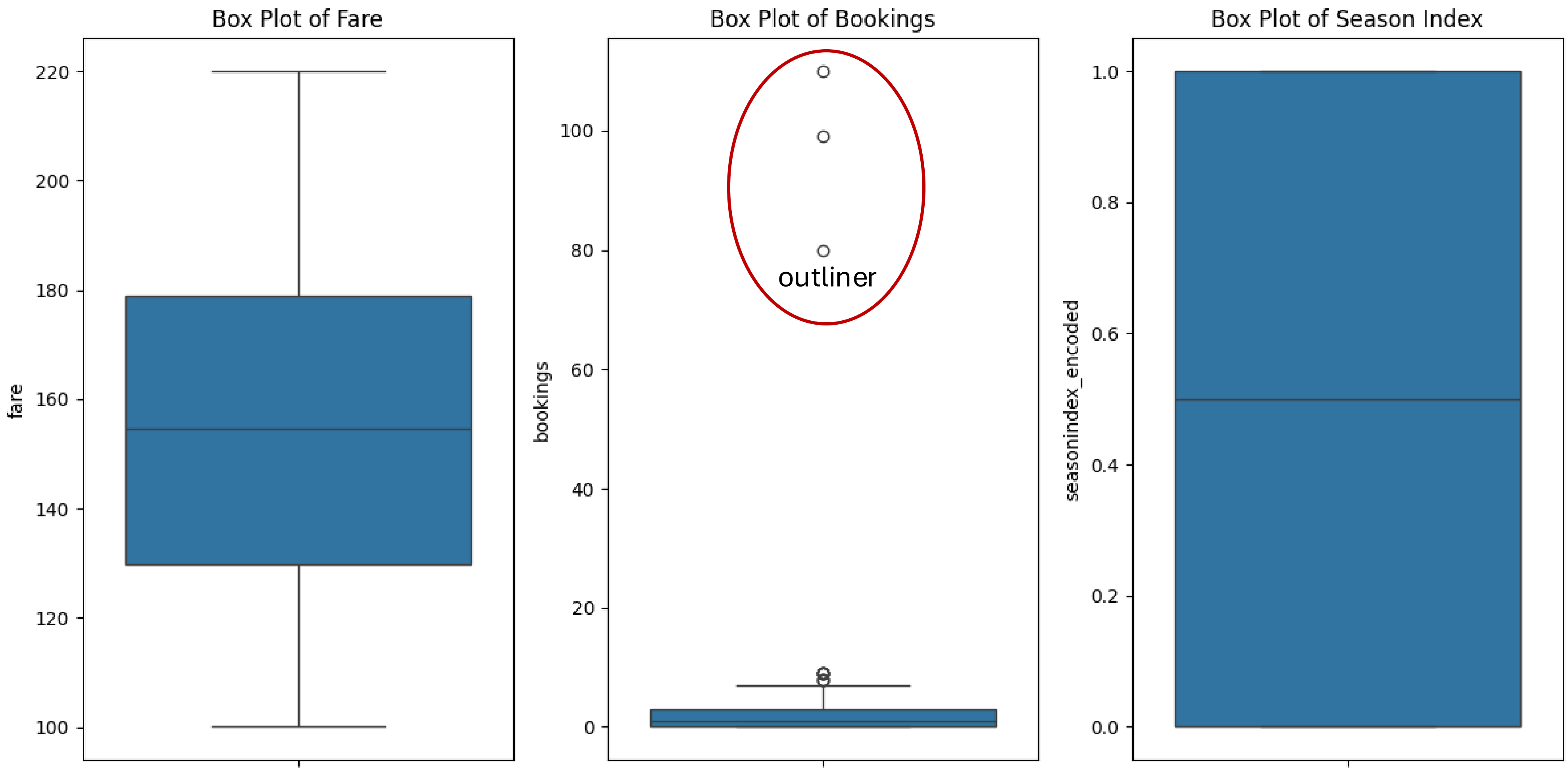
Exploratory data analysis (EDA)

	fare	bookings
count	1004.000000	1003.0000
mean	155.056375	1.953141
std	30.881664	5.470552
min	100.080000	0.000000
25%	129.737500	0.000000
50% (median)	154.650000	1.000000
75%	178.992500	3.000000
max	220.000000	110.000000
mode	N/A	0

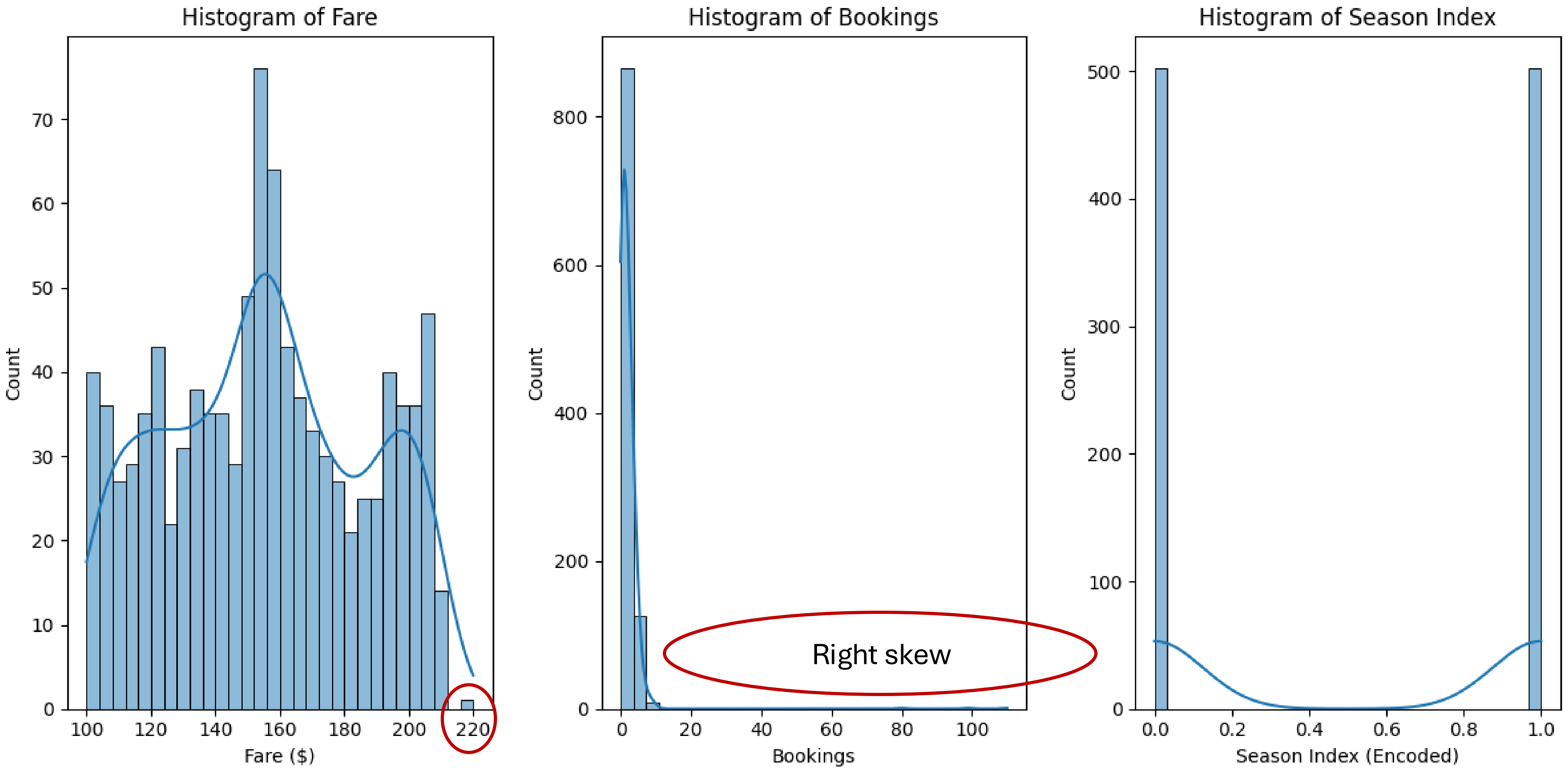
Missing data count, fare: 0 , bookings:1, seasonindex : 0

*The main purpose of EDA is to help look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events.

Exploratory data analysis (EDA)

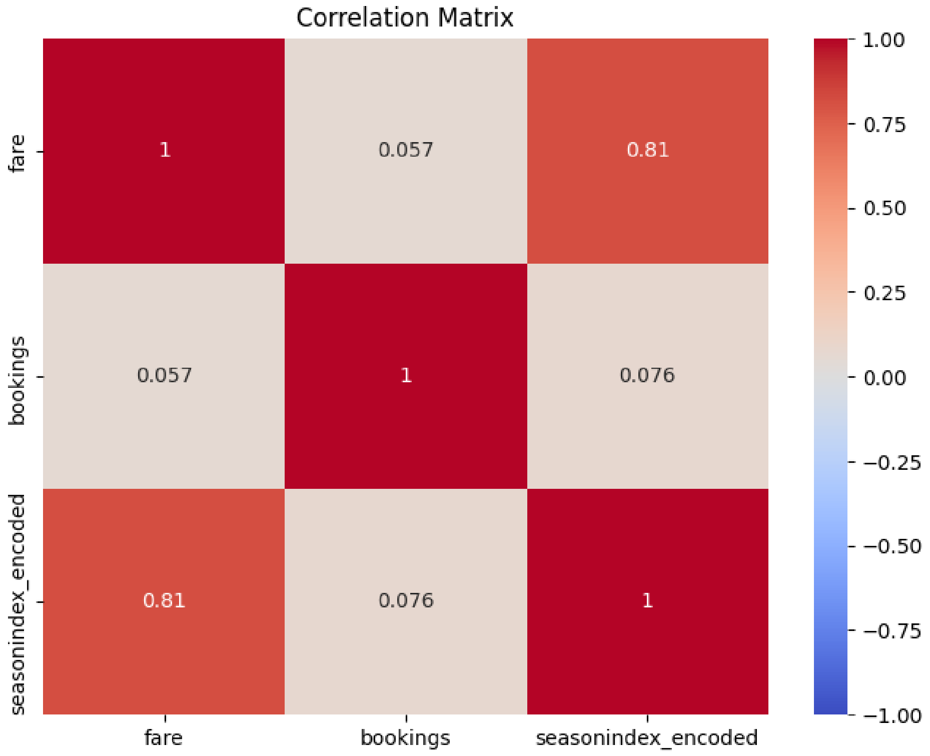


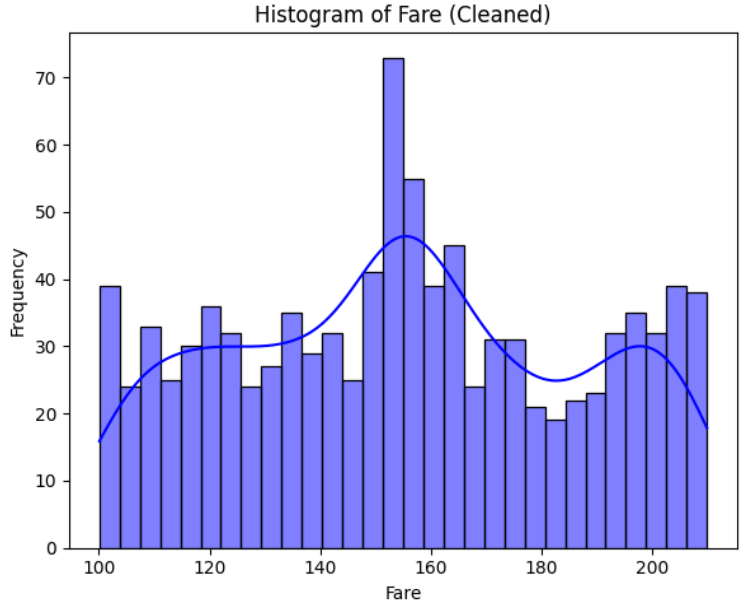
Exploratory data analysis (EDA)

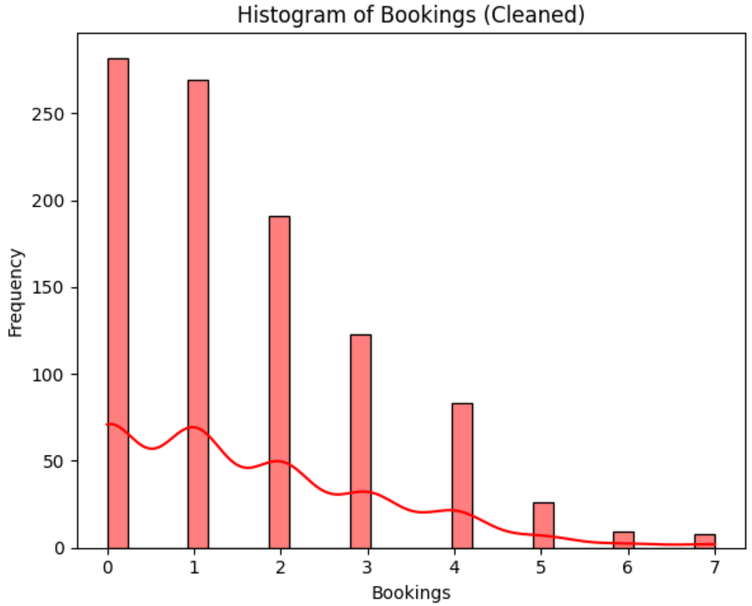


Exploratory data analysis (EDA): after remove outliers

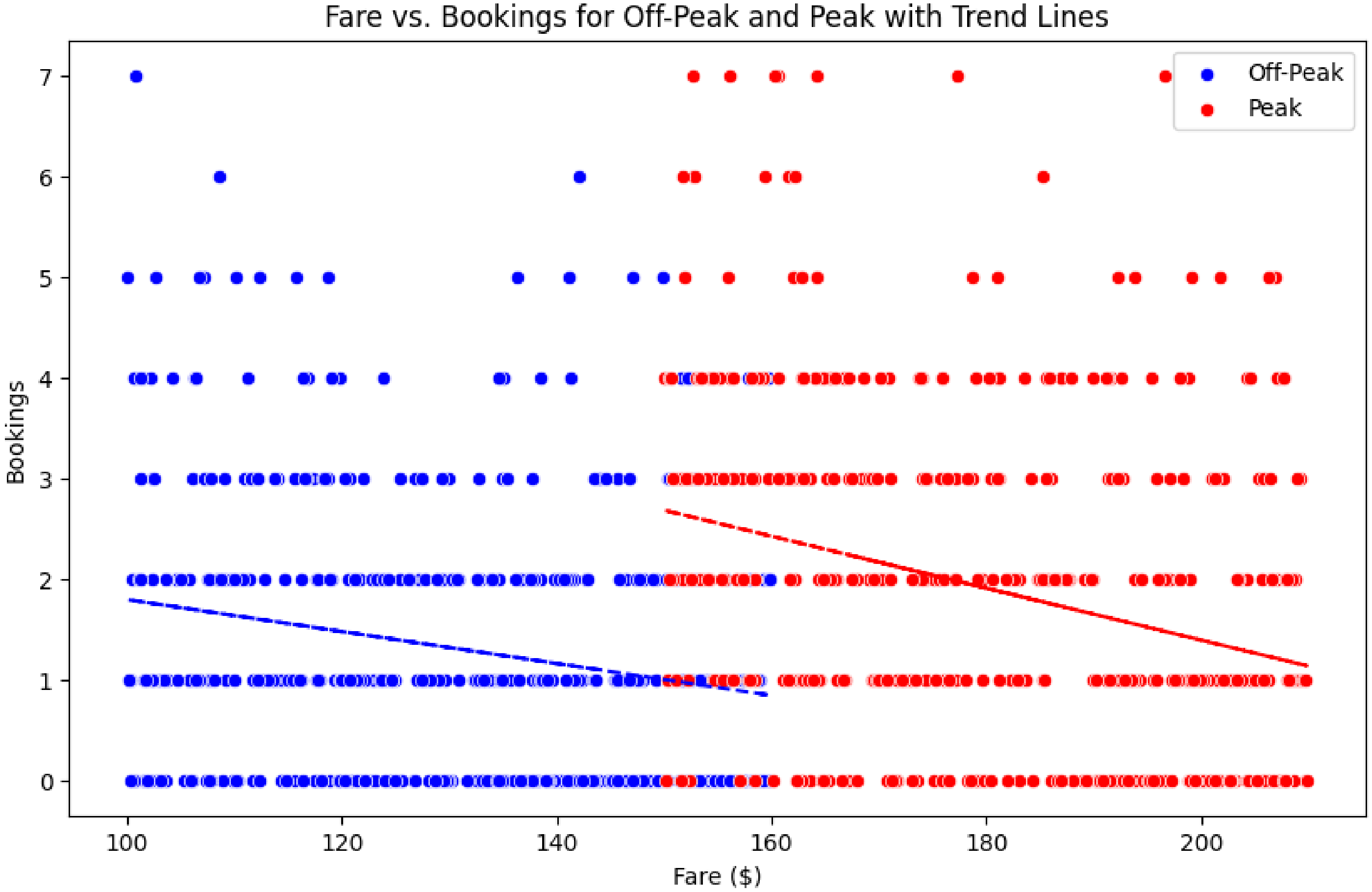
* A commonly used rule says that a data point is an **outlier** if it is more than 1.5 * {IQR} above the third quartile or below the first quartile.







Peak and Off- Peak



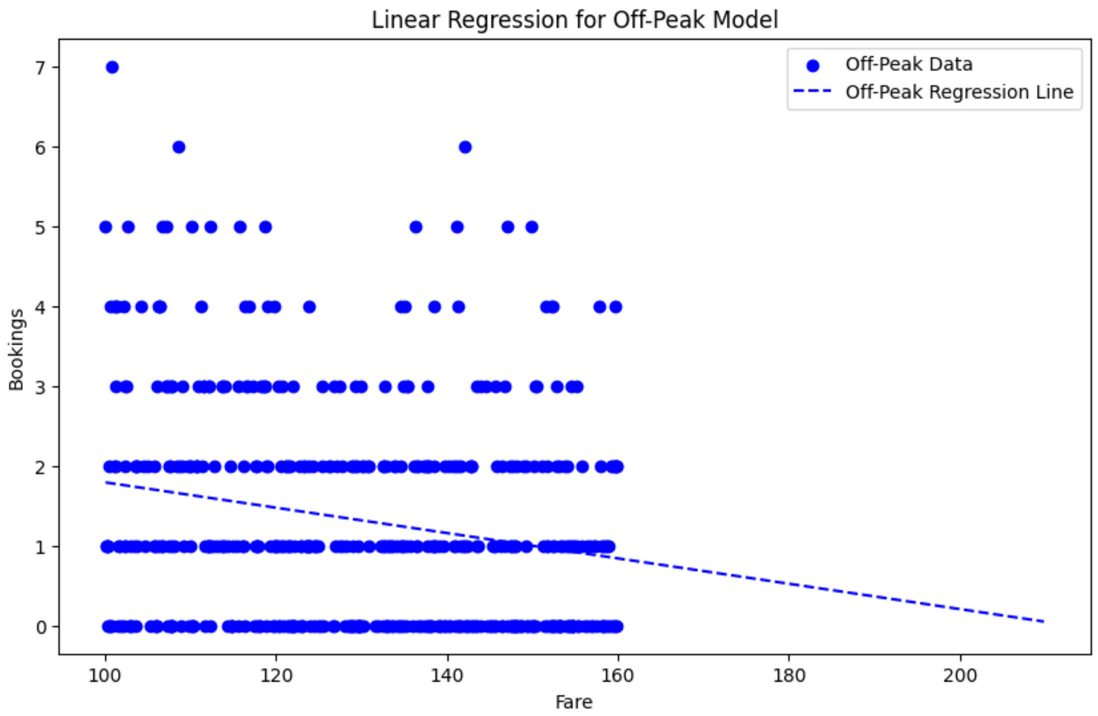
Independent t-test:

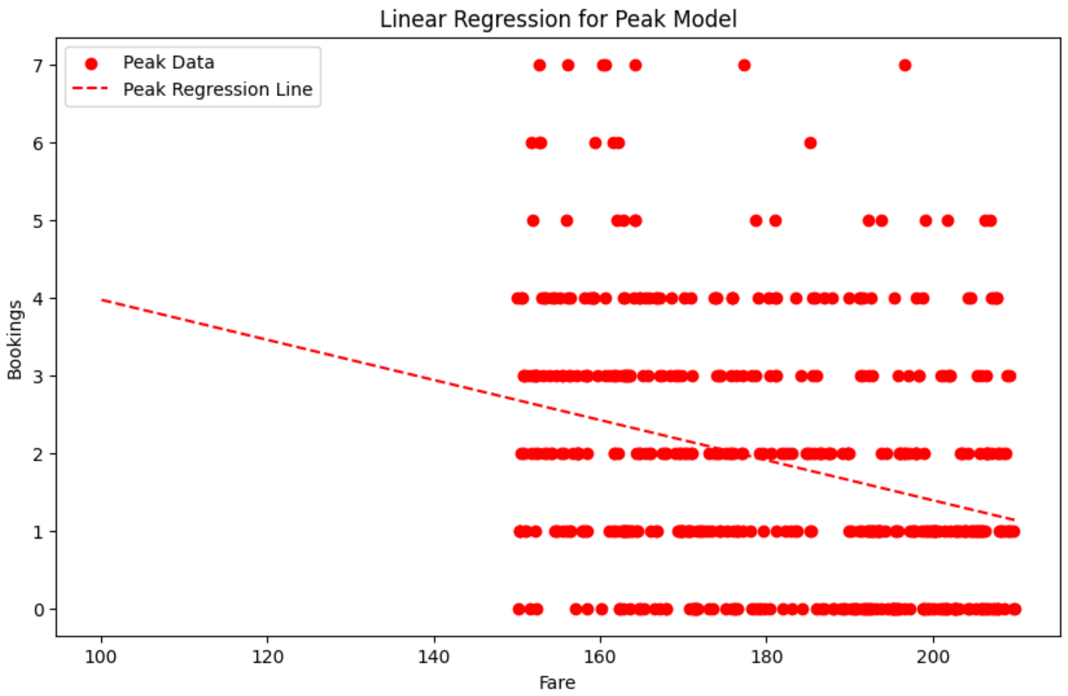
P-value: 0.0000000011590472660439186

<u>The relationship between fare and booking is likely to differ because customers may respond to high fares during peak season differently than to lower fares during the off-season.</u>

Model 1: Linear Regression

Bookings = α + β*Fare

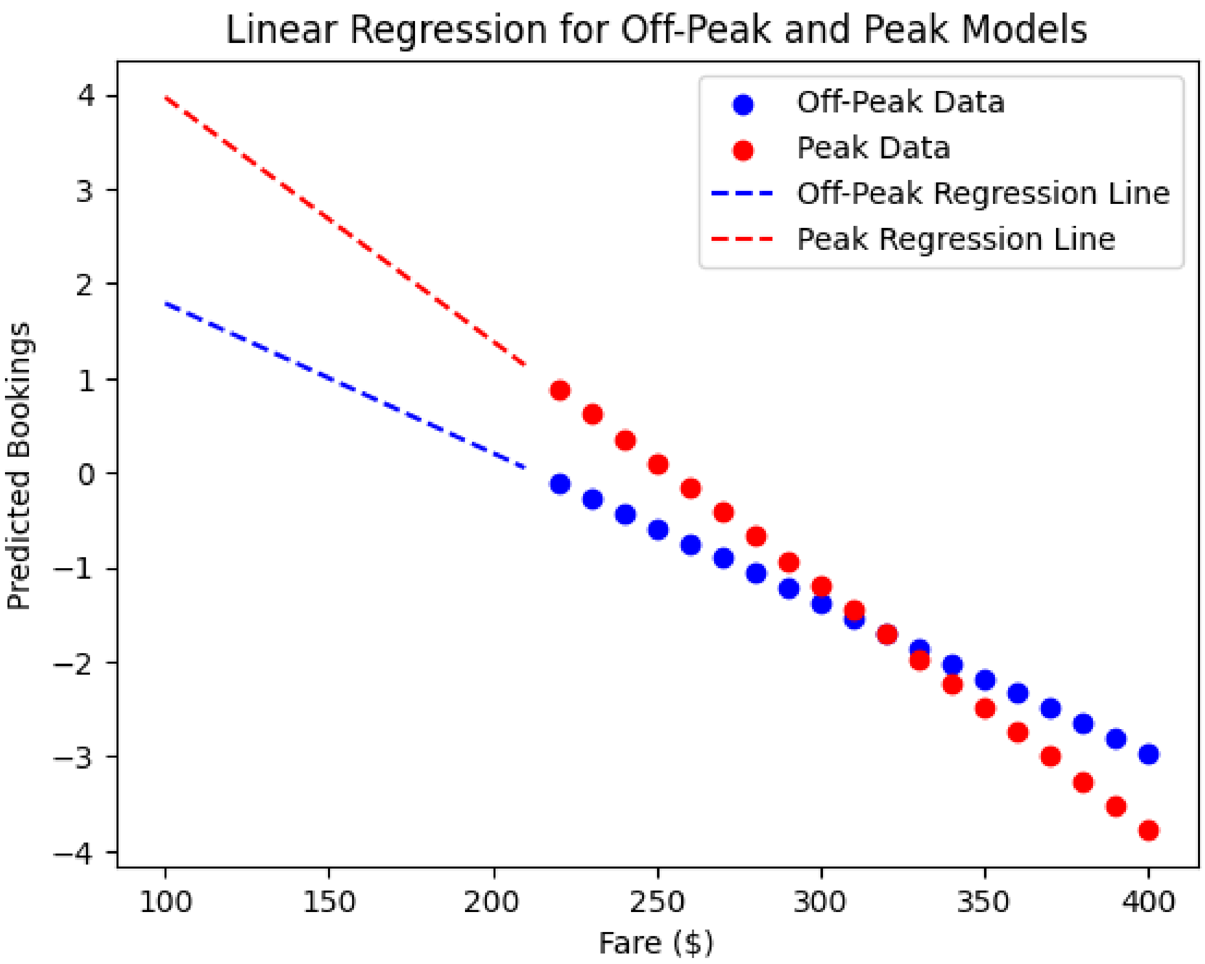




- Off-Peak Model: bookings = -0.015864051923704884 * fare + 3.3791320099290063
- Peak Model: bookings = -0.025809503747832304 * fare + 6.551925278933437
- R^2 for Off-Peak Model: 0.04244650475611633
- R^2 for Peak Model: 0.0851577460234264
- *R-squared (R2) is defined as a number that tells you how well the independent variable(s) in a statistical model explains the variation in the dependent variable. It ranges from 0 to 1, where 1 indicates a perfect fit of the model to the data.

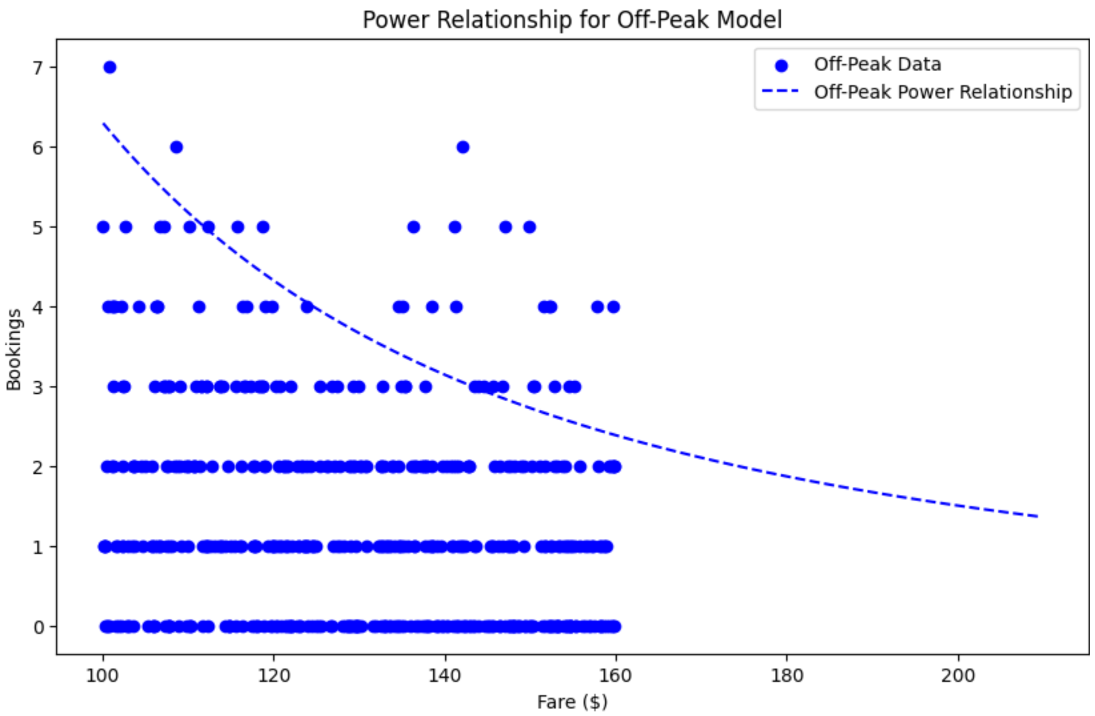
Model 1: Linear Regression - Prediction

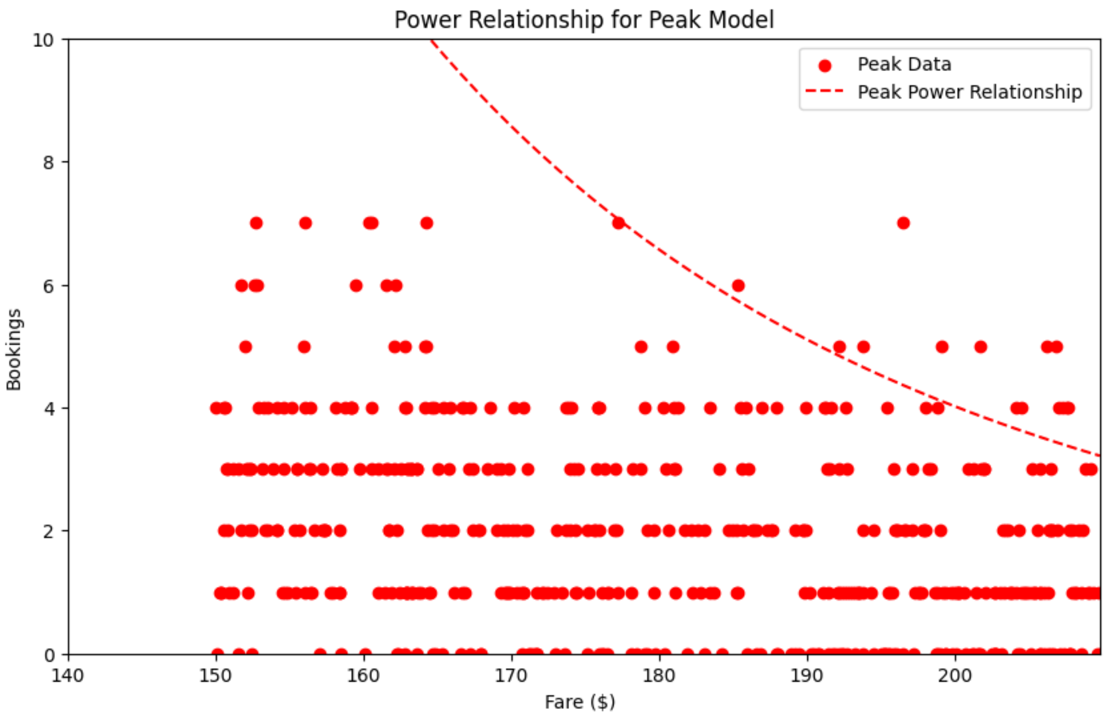
	fare	predicted_bookings_off_peak	predicted_bookings_peak
0	220	-0.110959	0.873834
1	230	-0.269600	0.615739
2	240	-0.428240	0.357644
3	250	-0.586881	0.099549
4	260	-0.745521	-0.158546
5	270	-0.904162	-0.416641
6	280	-1.062803	-0.674736
7	290	-1.221443	-0.932831
8	300	-1.380084	-1.190926
9	310	-1.538724	-1.449021
10	320	-1.697365	-1.707116
11	330	-1.856005	-1.965211
12	340	-2.014646	-2.223306
13	350	-2.173286	-2.481401
14	360	-2.331927	-2.739496
15	370	-2.490567	-2.997591
16	380	-2.649208	-3.255686
17	390	-2.807848	-3.513781
18	400	-2.966489	-3.771876



Model 2: Power

Bookings = $\alpha \text{Fare}^{\beta}$

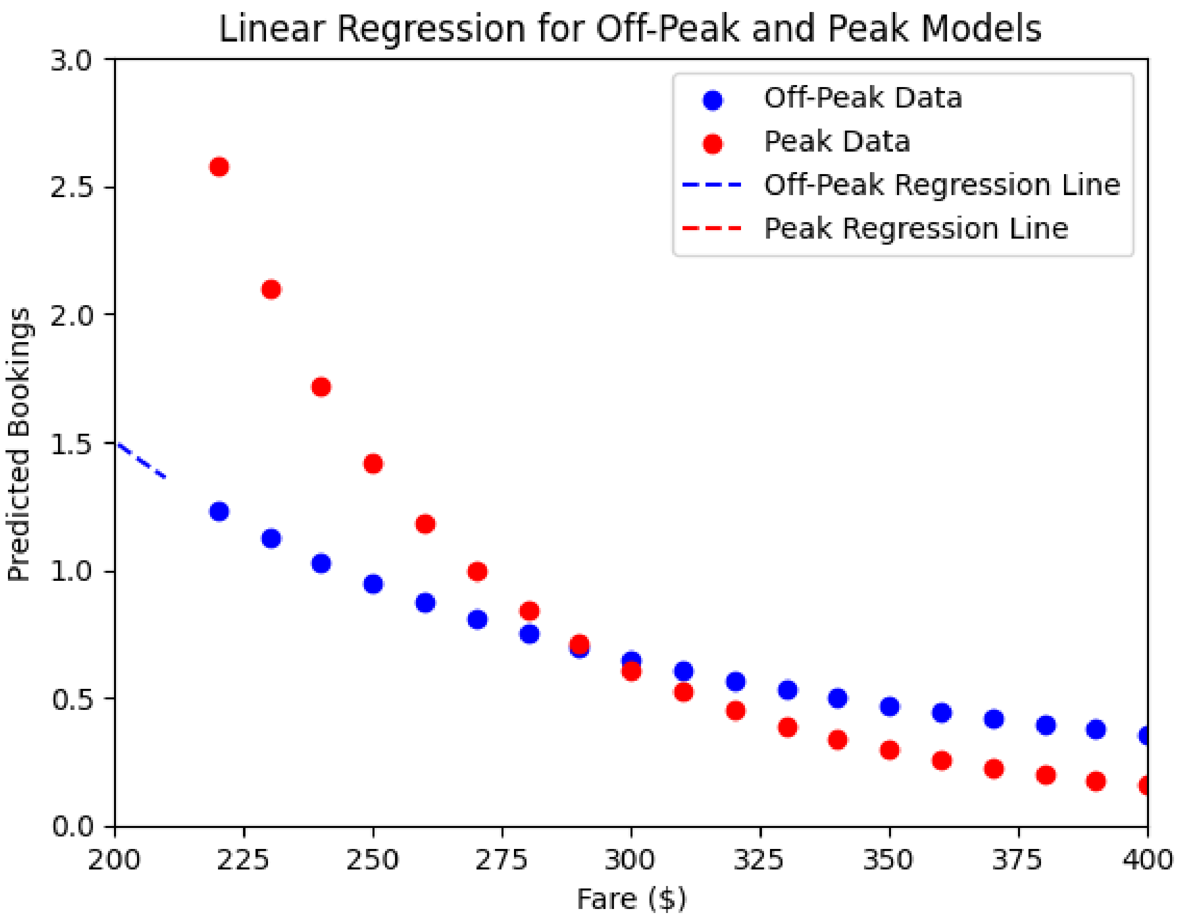




- Off-Peak Model: Bookings = 86152.30014202962 * Fare^-2.068050604983265
- Peak Model: Bookings = 211665339525.68036 * Fare^-4.659398892551365
- R^2 for Off-Peak Model: 0.043885844916571215
- R^2 for Peak Model: 0.08654025061422632

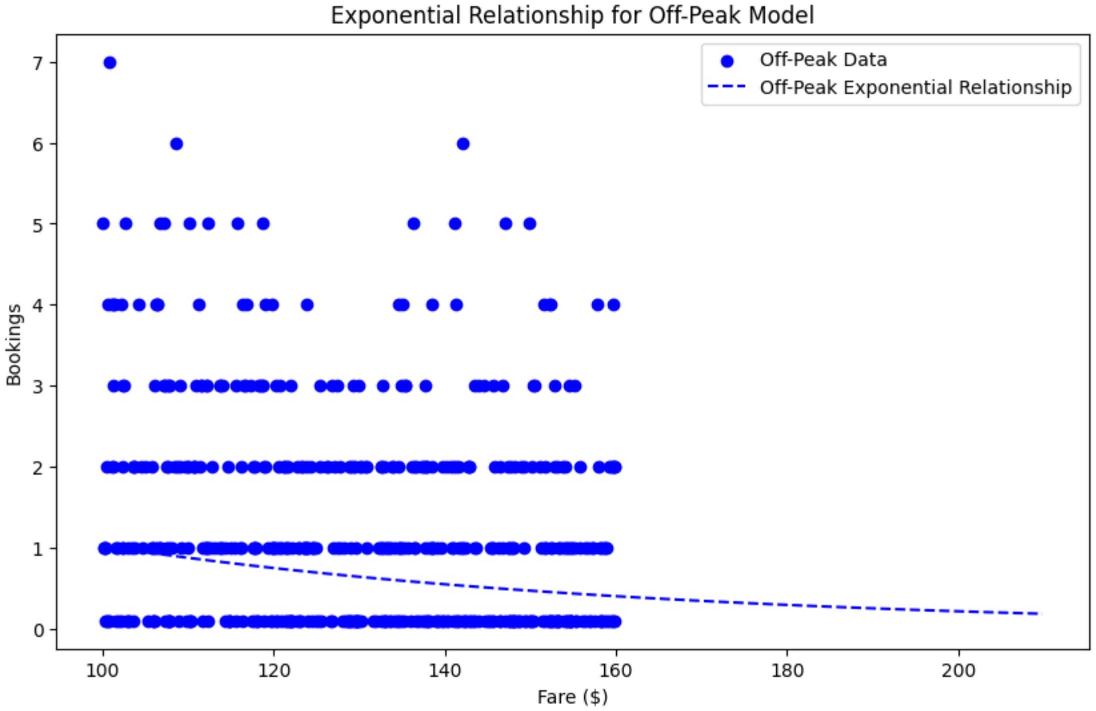
Model 2: Power - Prediction

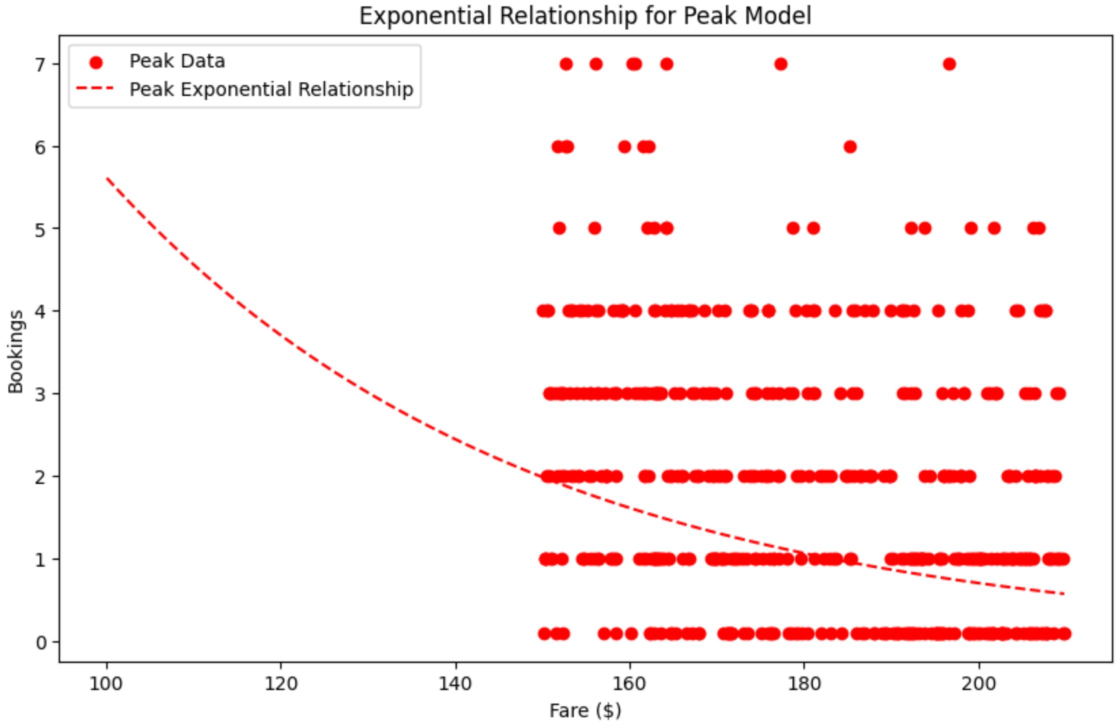
	fare	predicted_bookings_off_peak_log	predicted_bookings_peak_log
0	220	1.233157	2.578506
1	230	1.124849	2.096127
2	240	1.030077	1.719080
3	250	0.946685	1.421316
4	260	0.872931	1.183929
5	270	0.807390	0.993016
6	280	0.748894	0.838232
7	290	0.696471	0.711796
8	300	0.649314	0.607790
9	310	0.606743	0.521677
10	320	0.568185	0.449942
11	330	0.533154	0.389842
12	340	0.501234	0.339219
13	350	0.472069	0.296362
14	360	0.445353	0.259906
15	370	0.420819	0.228757
16	380	0.398239	0.202027
17	390	0.377411	0.178997
18	400	0.358158	0.159080



Model 3: Exponential

Bookings = α exp(β*Fare)

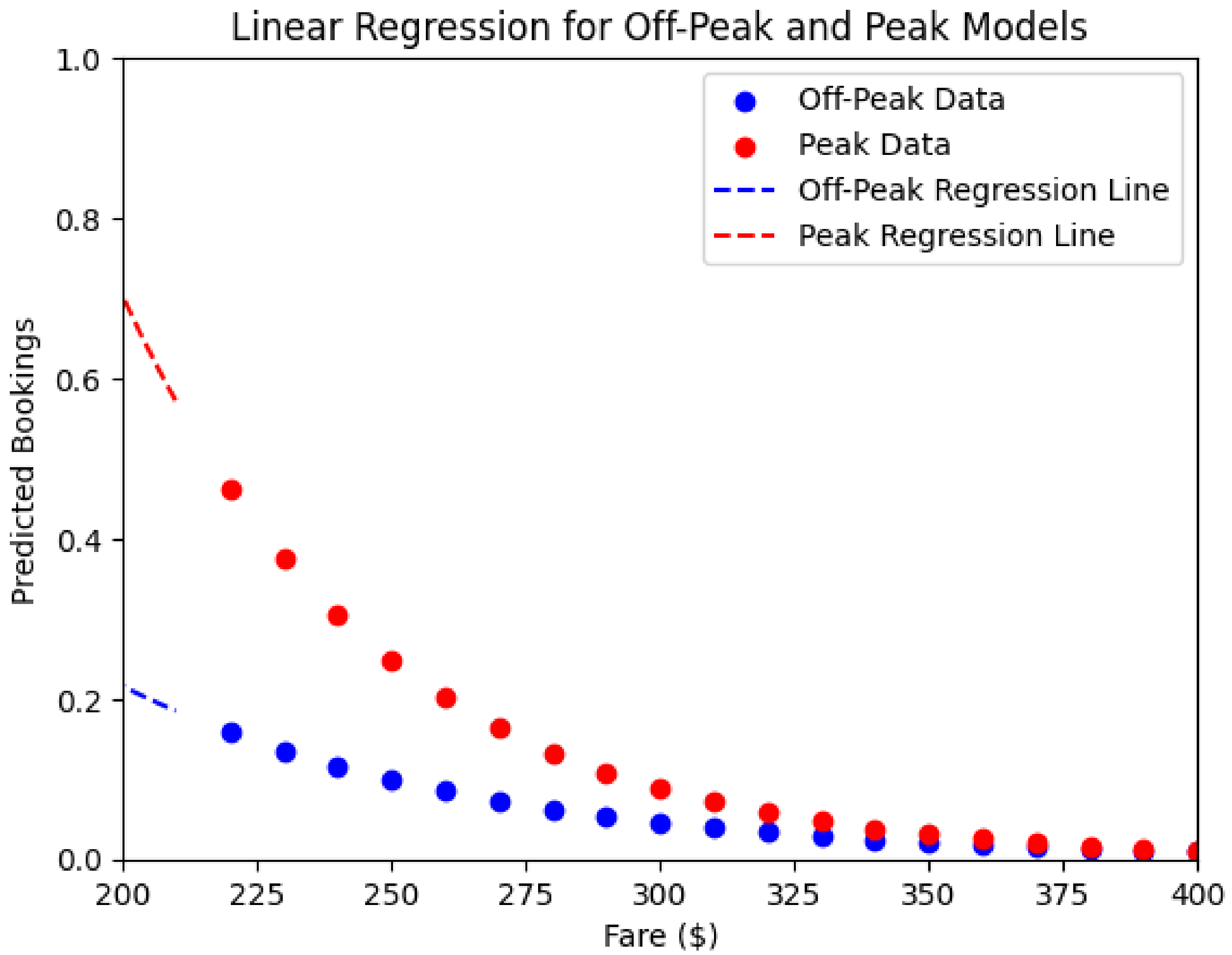




- Off-Peak Model: Bookings = 4.859565851583752 * exp(-0.015574838655681048 * Fare)
- Peak Model: Bookings = 44.94380135302284 * exp(-0.02079634676039662 * Fare)
- R^2 for Off-Peak Model: 0.036637104718004854
- R^2 for Peak Model: 0.07836908166721535

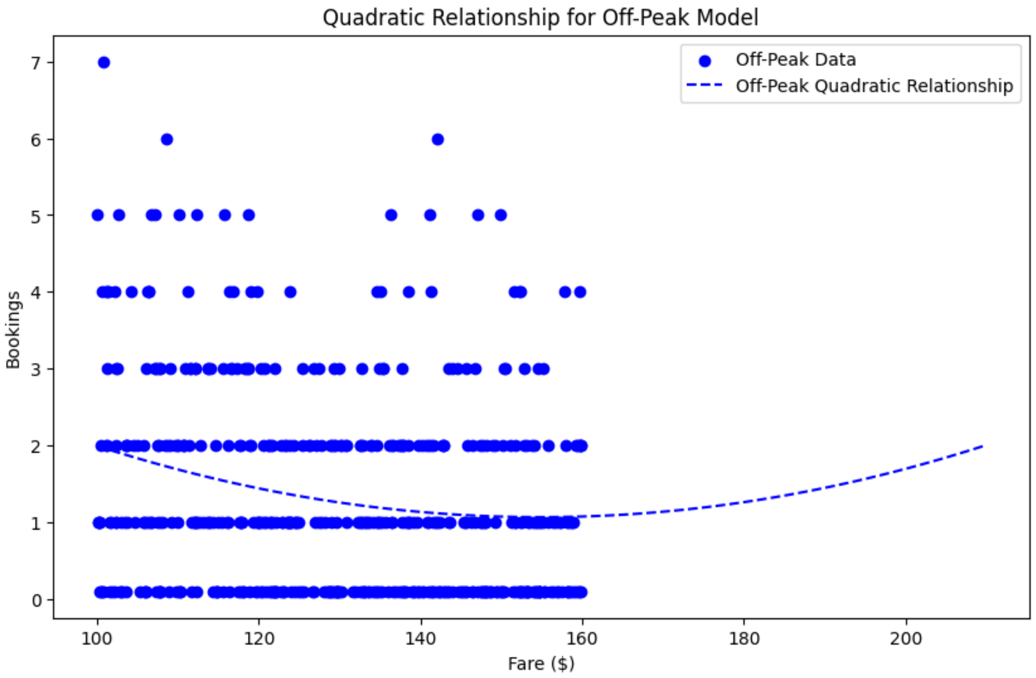
Model 3: Exponential - Prediction

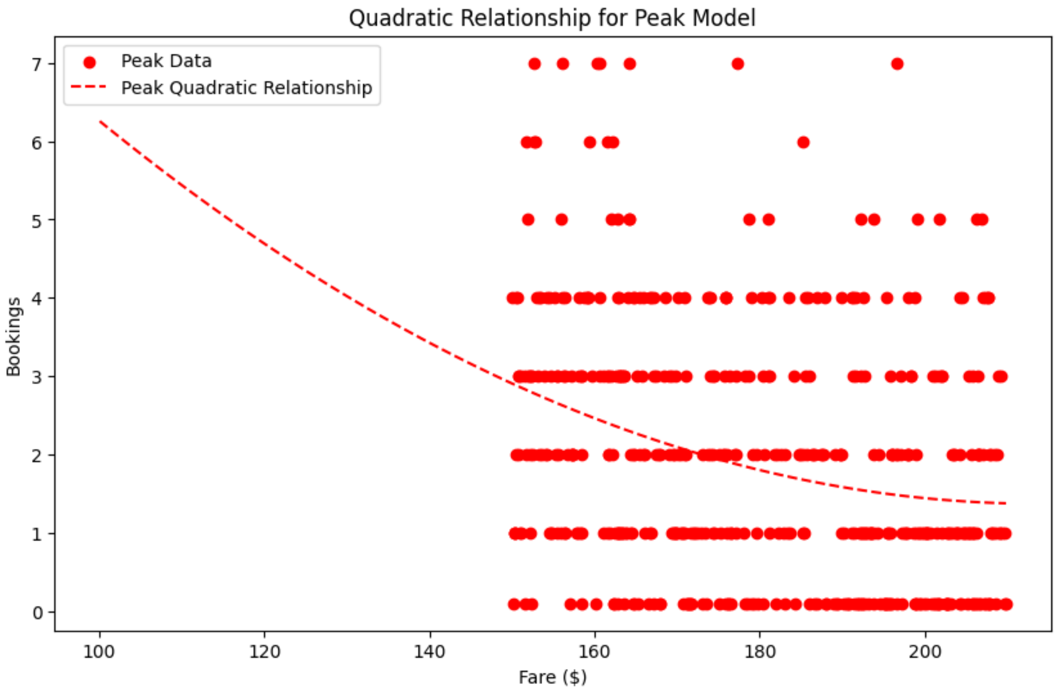
	fare	predicted_bookings_off_peak_exp	predicted_bookings_peak_exp
0	220	0.157944	0.463113
1	230	0.135164	0.376158
2	240	0.115670	0.305529
3	250	0.098988	0.248162
4	260	0.084711	0.201566
5	270	0.072494	0.163719
6	280	0.062038	0.132979
7	290	0.053091	0.108010
8	300	0.045434	0.087730
9	310	0.038881	0.071258
10	320	0.033273	0.057878
11	330	0.028474	0.047011
12	340	0.024368	0.038184
13	350	0.020853	0.031014
14	360	0.017846	0.025191
15	370	0.015272	0.020461
16	380	0.013069	0.016619
17	390	0.011184	0.013499
18	400	0.009571	0.010964



Model 4: Quadratic

Bookings = α + β*Fare + γ*Fare^2

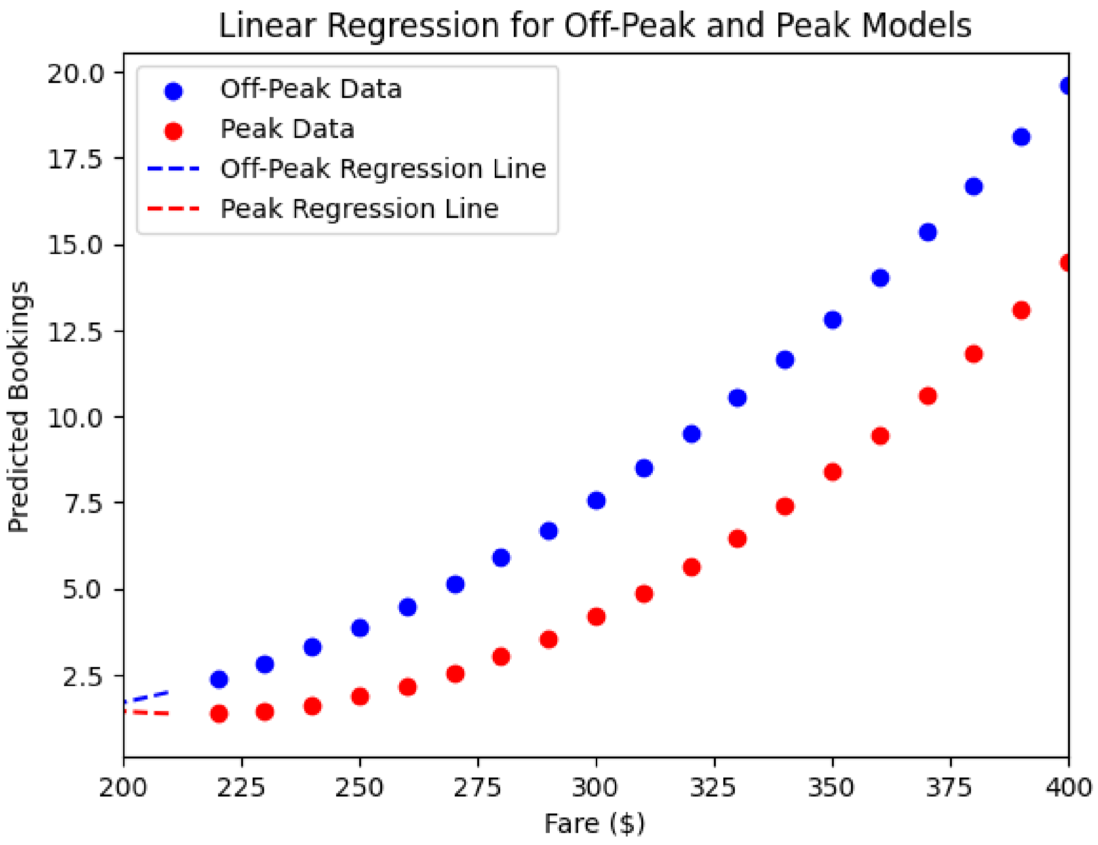




- Off-Peak Model: Bookings = 8.471194356166503 + -0.0956495513970951 * Fare + 0.00030885350764885914 * Fare^2
- Peak Model: Bookings = 18.64014691451785 + -0.16157659448724912 * Fare + 0.00037796591016549843 * Fare^2
- R^2 for Off-Peak Model: 0.04631332094060647
- R^2 for Peak Model: 0.08864035312259388

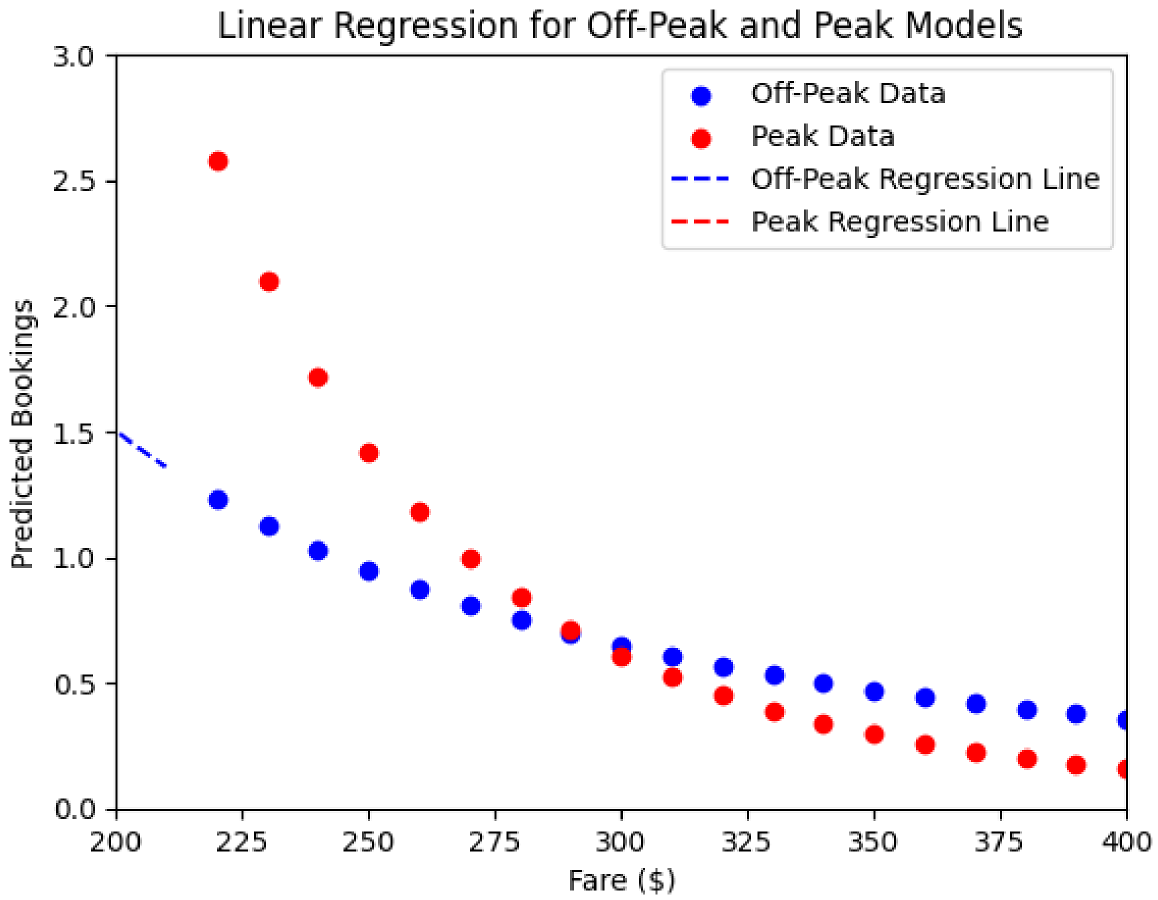
Model 4: Quadratic - Prediction

	fare	predicted_bookings_off_peak_exp	predicted_bookings_peak_exp
0	220	2.376803	1.386846
1	230	2.810148	1.471927
2	240	3.305264	1.632601
3	250	3.862151	1.868868
4	260	4.480808	2.180728
5	270	5.161236	2.568181
6	280	5.903435	3.031228
7	290	6.707404	3.569868
8	300	7.573145	4.184100
9	310	8.500656	4.873927
10	320	9.489937	5.639346
11	330	10.540989	6.480358
12	340	11.653812	7.396964
13	350	12.828406	8.389163
14	360	14.064770	9.456955
15	370	15.362906	10.600340
16	380	16.722811	11.819318
17	390	18.144488	13.113890
18	400	19.627935	14.484055



Final Model - Model 2: Power

- Data uncertainty: need to think about how to address the outlier (irregular booking behavior) => may need another model
- Need more booking data for fare > $300
- R^2 is small, lack explainability
- Conditional machine learning model may work better (XGBoost)



Thank you for your attention.