

1.

- a. Before: 162833
After: 157431
- b. Dep airport and Arr airport both had NaN values
- c. 32 unique Dep-Arr formats

...		Travel Date	Dep. airport	Dep. time	Arr. airport	Arr. time	Duration	Direct	Airfare(NZ\$)	Arrival and Departure
	4	2019-09-19	AKL	9:00 AM	CHC	10:25 AM	1h 25m	(Direct)	133	CHC - AKL
	7	2019-09-19	AKL	11:00 AM	CHC	12:25 PM	1h 25m	(Direct)	163	CHC - AKL
	8	2019-09-19	AKL	7:00 PM	CHC	8:25 PM	1h 25m	(Direct)	163	CHC - AKL
	10	2019-09-19	AKL	10:00 AM	CHC	11:25 AM	1h 25m	(Direct)	193	CHC - AKL
	11	2019-09-19	AKL	4:00 PM	CHC	5:25 PM	1h 25m	(Direct)	193	CHC - AKL

d.

</>		Travel Date	Dep. airport	Dep. time	Arr. airport	Arr. time	Duration	Direct	Airfare(NZ\$)	Arrival and Departure
	4	2019-09-19	AKL	9	CHC	10	85	0	133	CHC - AKL
	7	2019-09-19	AKL	11	CHC	12	85	0	163	CHC - AKL
	8	2019-09-19	AKL	19	CHC	20	85	0	163	CHC - AKL
	10	2019-09-19	AKL	10	CHC	11	85	0	193	CHC - AKL
	11	2019-09-19	AKL	16	CHC	17	85	0	193	CHC - AKL

e.

2.

- a. Xtrain:4181, len: 118055
Ytrain:4181 len: 118055
Xtest:138367 len: 39352
YTest:138367 len: 39352
- b. R^2 : 0.27370029784417715 (not the best fit, low correlation, closer to 1 would be a better fit)
B1, B2: [-1.00881225e-01, 1.96070407e+02]
B0: 260.37139390206437
- c. 108.16120456108199 (MAE), given the fit was not good a mean absolute error as high as this should be expected