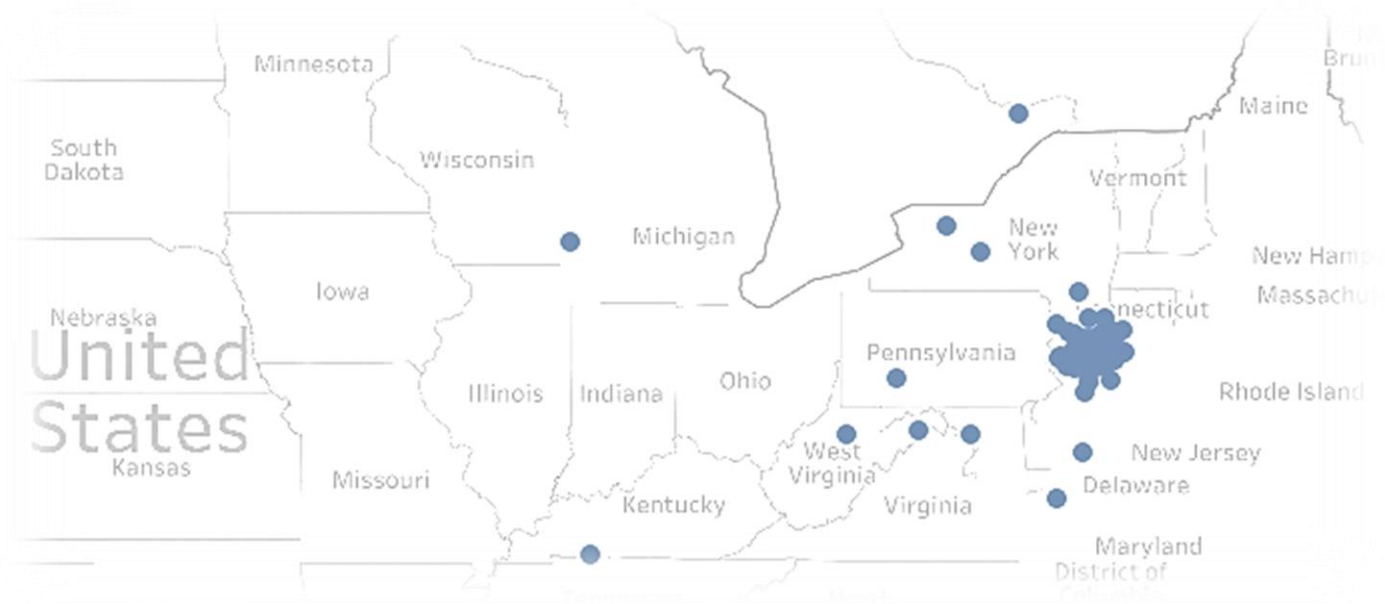
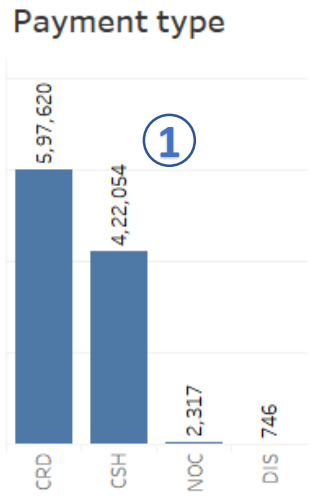
Trip Fare Prediction

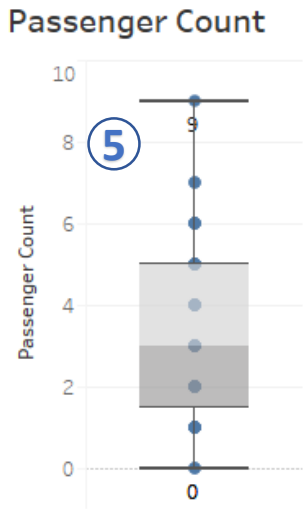
Foreword

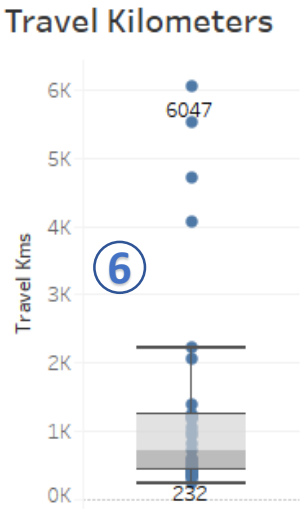
- To predict the trip fare amount using the passengers travel points and time.
- Rate category – the dominant factor for fare estimation

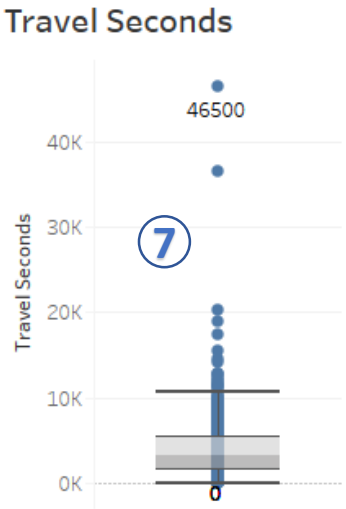


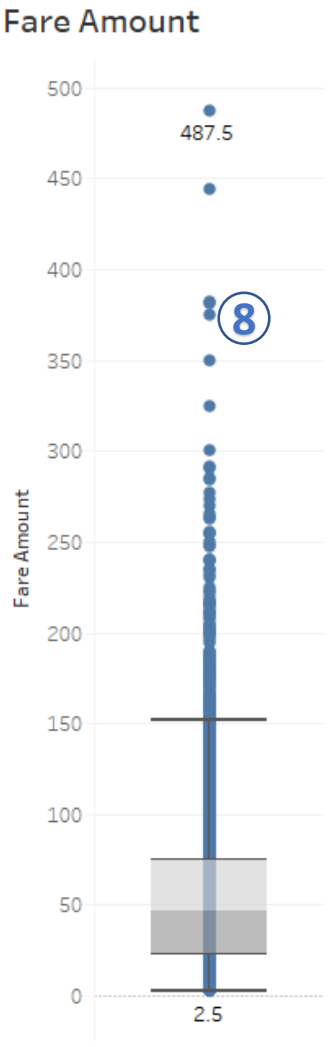
Data Overview

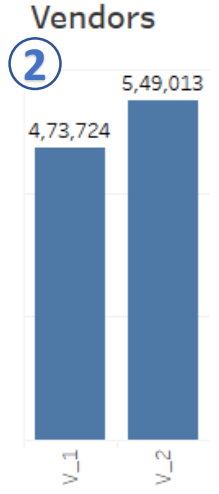


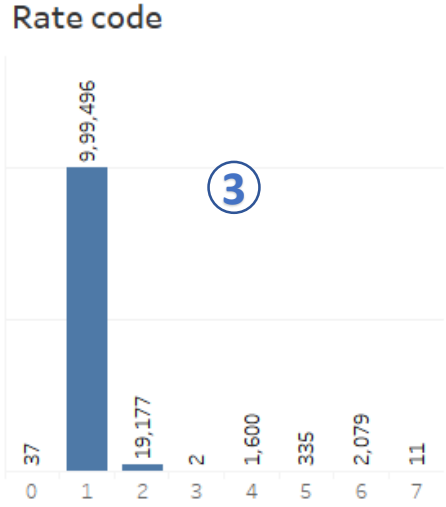


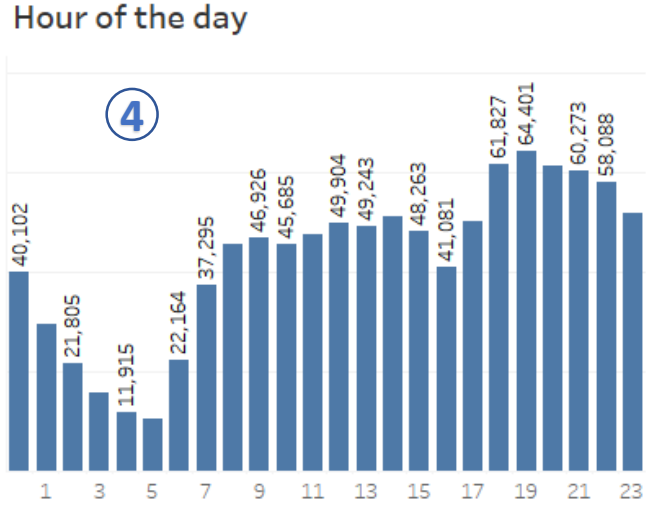






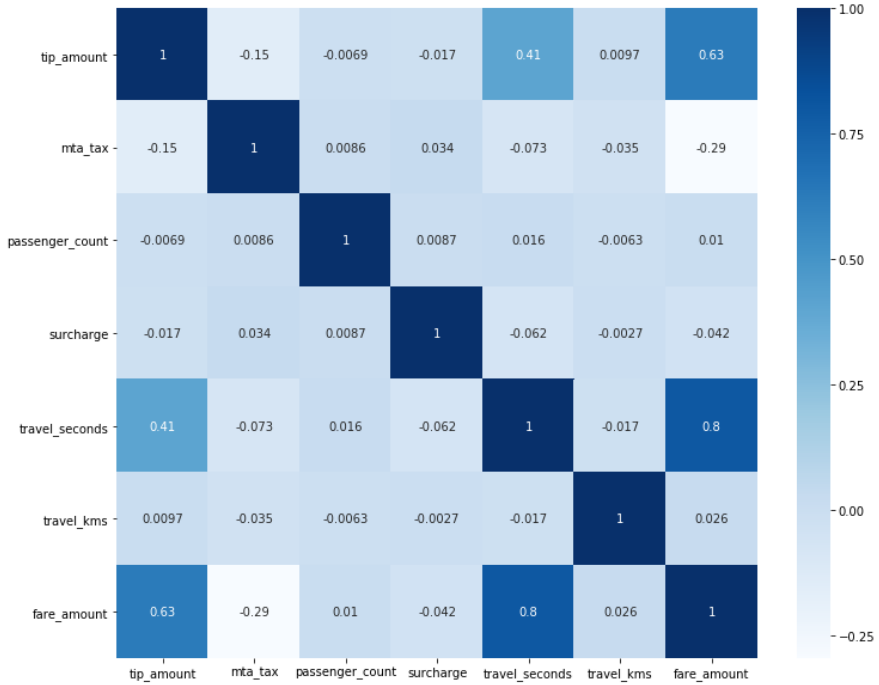


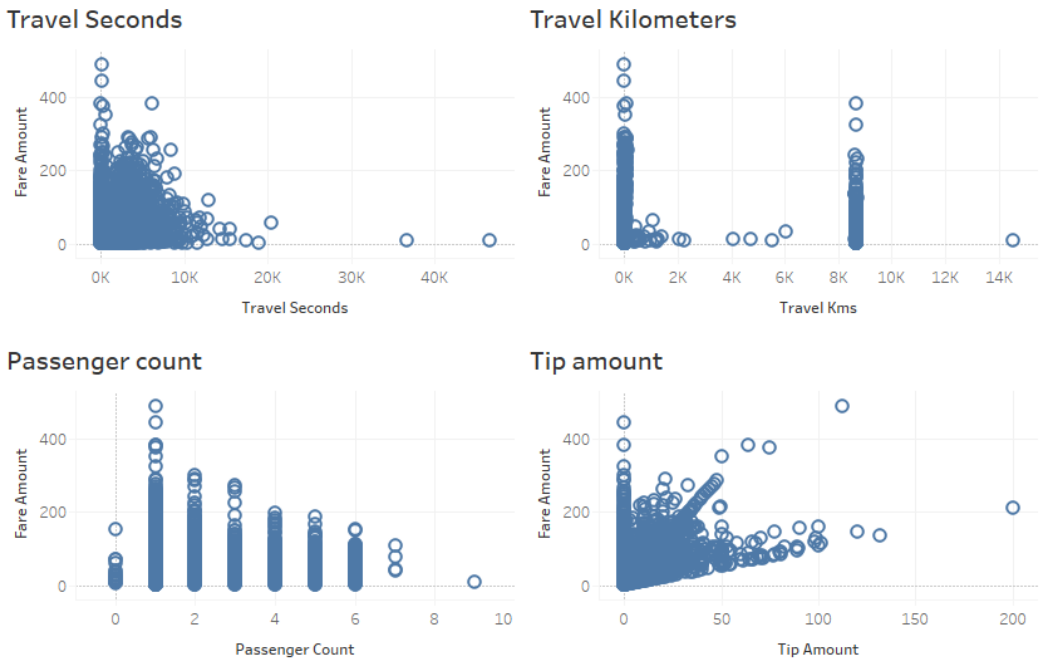




1) Payment method opted
2) Trip providers (Vendors)
3) Rate Category (Code)
4) Time of Departure
5) Number of passengers
6) Travel distance
7) Travel duration
8) Trip fare (Amount)

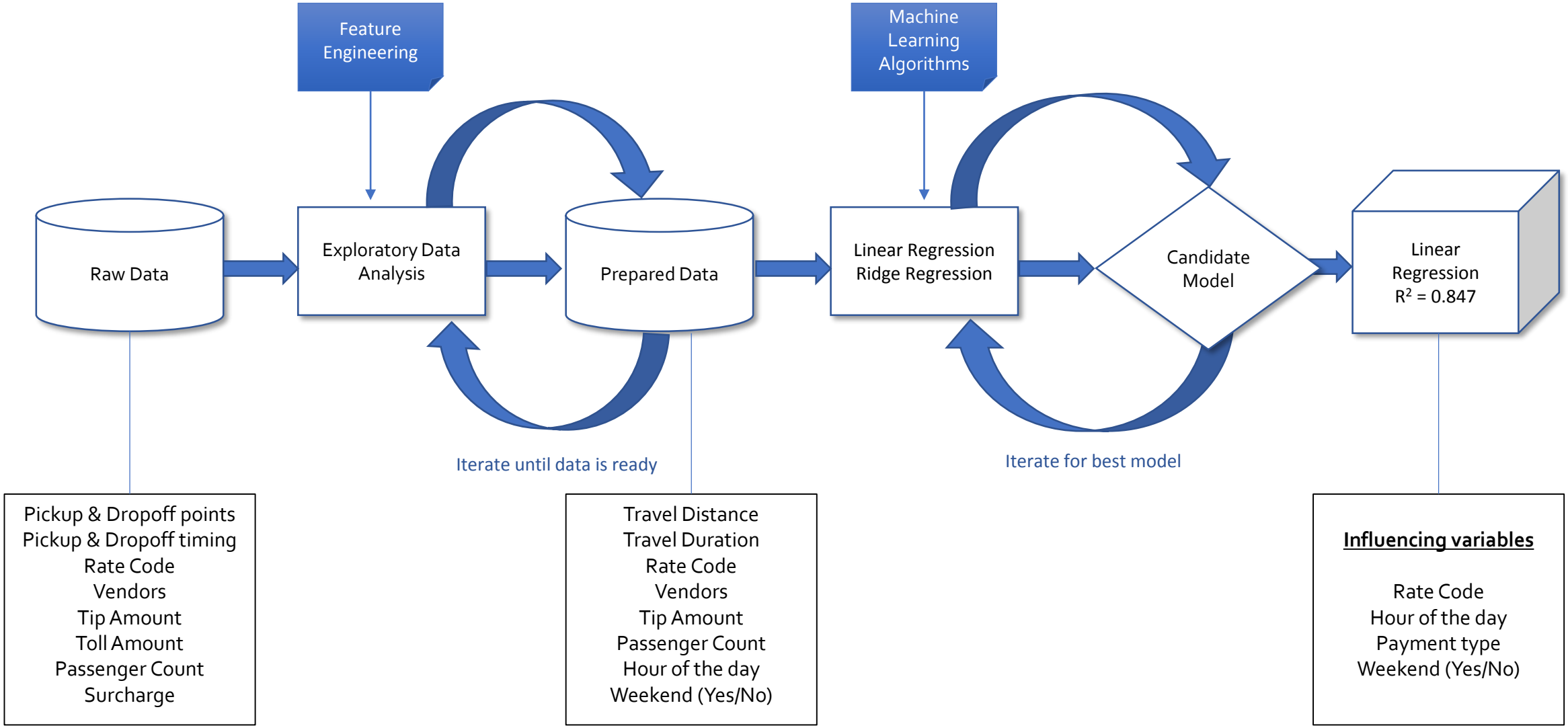
Data Preparation



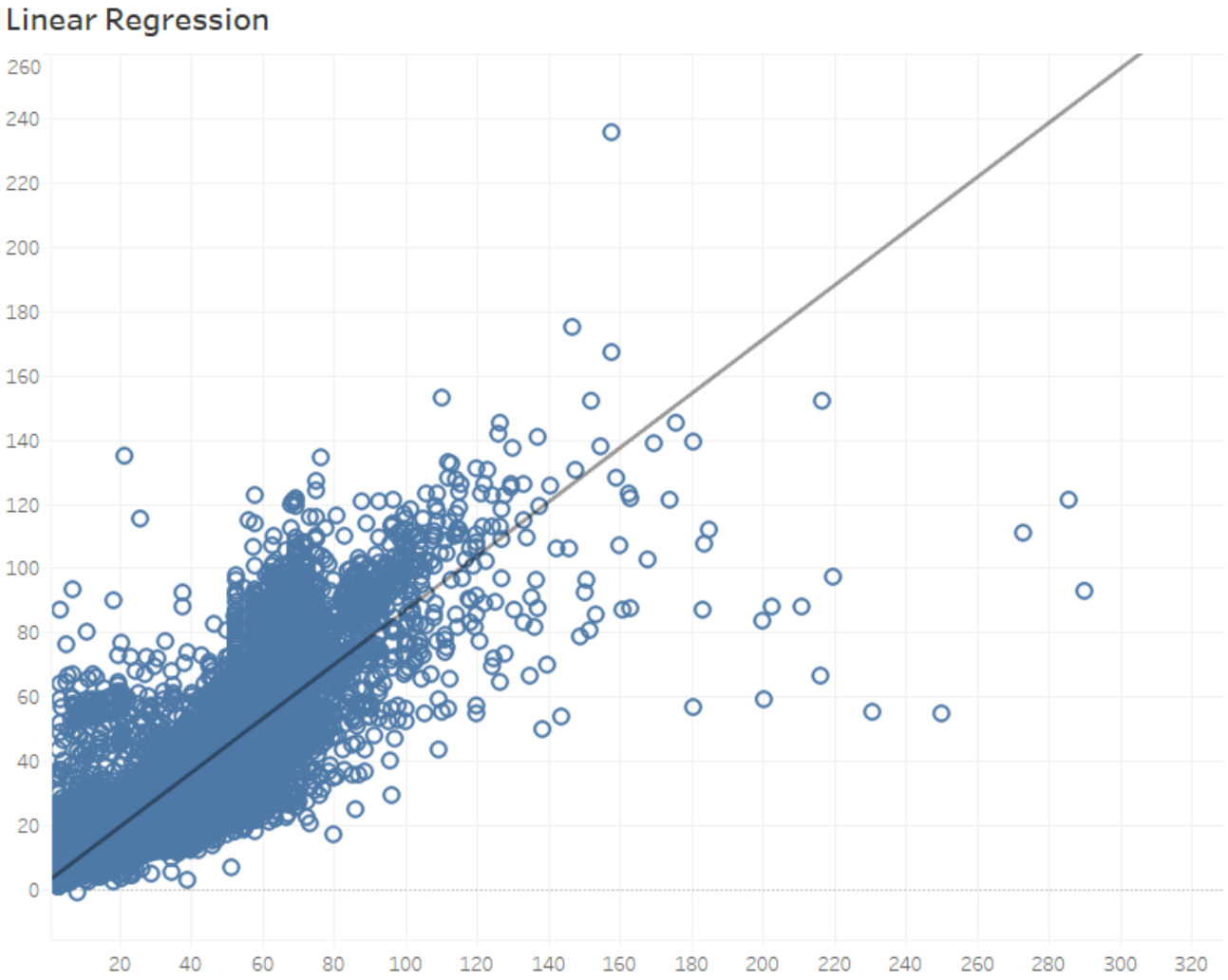


	tip_amount	passenger_count	fare_amount	travel_seconds	travel_kms	vendors	payment_type	rate_code	hour	is_weekend
0	1.4	1	8.4	360.0	1.311173	DST000401	CRD	1	4	0
1	1.0	3	8.5	360.0	2.596270	DST000401	CRD	1	18	1
2	0.0	2	7.0	360.0	1.538152	DST000401	CSH	1	8	0
3	1.8	2	11.3	720.0	1.598931	DST000532	CRD	1	9	0
4	0.0	1	10.0	840.0	1.626473	DST000401	CSH	1	13	0

The Learning Process



Model Summary



OLS Regression Results

Dep. Variable:	fare_amount	**R-squared:**	0.847
Model:	OLS	**Adj. R-squared:**	0.847
Method:	Least Squares	**F-statistic:**	1.453e+05
Date:	Sun, 27 Jan 2019	**Prob (F-statistic):**	0.00
Time:	16:55:29	**Log-Likelihood:**	-3.0698e+06
No. Observations:	1022737	**AIC:**	6.140e+06
Df Residuals:	1022697	**BIC:**	6.140e+06
Df Model:	39		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	14.9064	0.341	43.758	0.000	14.239	15.574

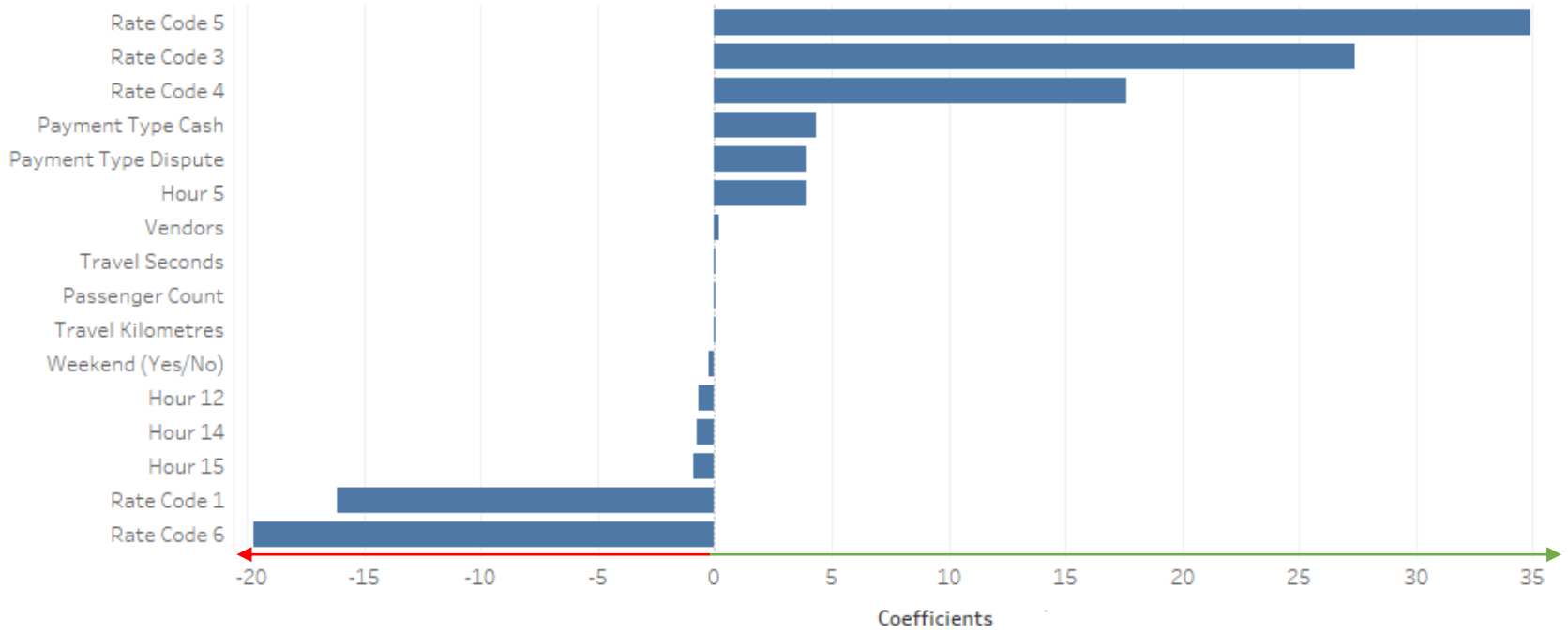
Omnibus:	621441.025	**Durbin-Watson:**	1.995
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	6594070031.115
Skew:	1.043	**Prob(JB):**	0.00
Kurtosis:	396.364	**Cond. No.**	1.26e+16

Best fit line

Coefficient of Determination: 0.847

Root Mean Square Error: 4.723

Influencing Features

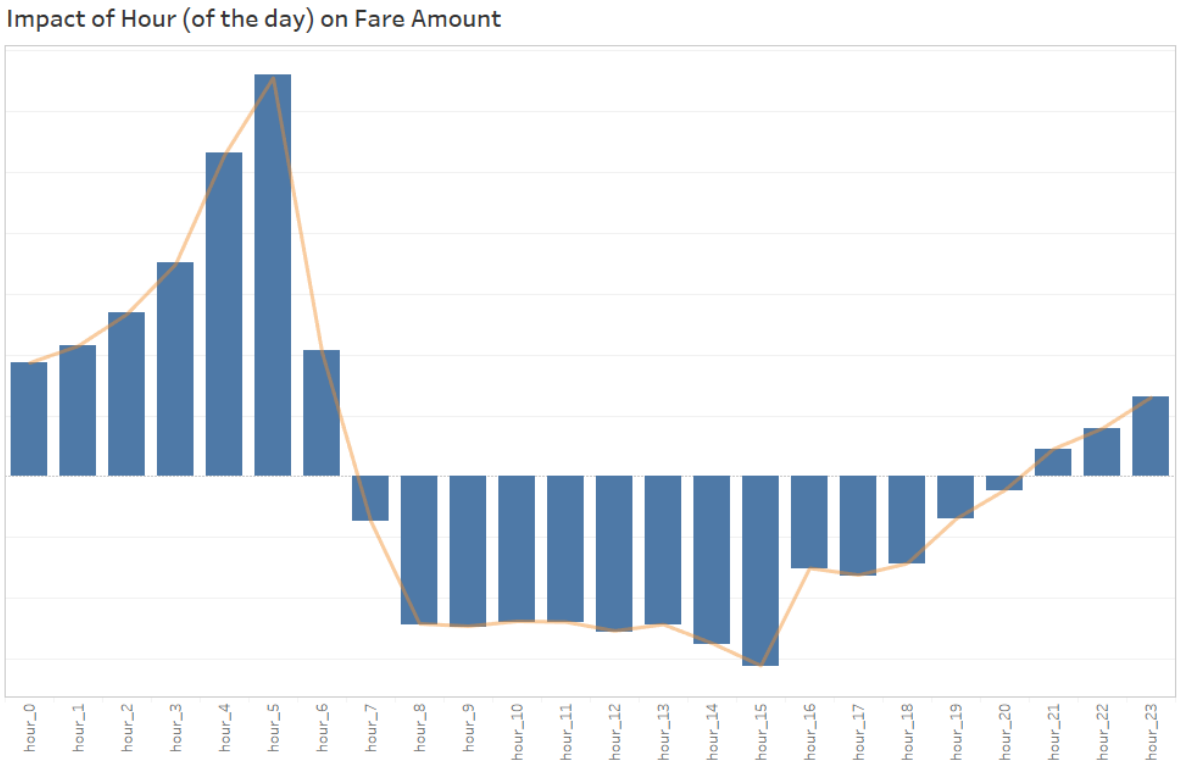


Intercept	14.906429
is_weekend[T.1]	-0.247151
tip_amount	1.721869
passenger_count	0.004502
travel_seconds	0.011879
travel_kms	0.000294
vendors	0.173938
payment_type_CRD	3.407051
payment_type_CSH	4.340676
payment_type_DIS	3.917499
payment_type_NOC	3.241203
rate_code_0	-14.651207
rate_code_1	-16.143845
rate_code_2	3.777963
rate_code_3	27.361996
rate_code_4	17.604424
rate_code_5	34.865875
rate_code_6	-19.751808
rate_code_7	-18.156970
hour_0	1.554615
hour_1	1.715662
hour_2	1.985119
hour_3	2.369083
hour_4	3.294756
hour_5	3.875275
hour_6	1.659390
hour_7	0.278456
hour_8	-0.610778
hour_9	-0.637590
hour_10	-0.567454
hour_11	-0.574707
hour_12	-0.660471
hour_13	-0.593586
hour_14	-0.780043
hour_15	-0.942915
hour_16	-0.135123
hour_17	-0.191727
hour_18	-0.091318
hour_19	0.265025
hour_20	0.514578
hour_21	0.856597
hour_22	1.038401
hour_23	1.285181

- Rate Category (Code) – based on travel points, passengers count etc.
- Hour of the day – start time of the journey
- Payment type – Cash (CSH), Card (CRD), Dispute (DIS), No Charge (NOC)
- Travel duration – time between start and end of journey in seconds
- Travel distance – distance between start and end points in kilometres
- Weekend (Yes/No) – 1 if the travel is made on weekend

Inference

- Fare amount is increasing during night hours and getting reduced during day.
- Reduction in fare amount during weekends to invite more customers.
- People have preferred card payment over cash because the fare amount for cash payment is slightly higher than the other.
- Rate code 1 (standard city rate) follows a nominal fare pattern. Rate code 2 and 3 (JFK and Newark airport) for which the fare amount is considerably high. Rate code 4 and 5 (out of city boundaries) for which metered rate is adopted. The difference being rate code 5 has automatic metering system.
- The number of people who opted for the service at 5 am is too low. This could be probably because the fare amount is too high at the point.



Discussions

- Fare reduction can also be given for travel on weekdays which could probably make the customer's prefer the company's travel option to a better extent.
- Reduction in the dynamic fare set for 5 am may possibly increase the number of people opting for the service at the time

Thank You