ABC Supply Hackathon

Team 8
Rishabh Joshi
Anisha Dubhashi
5/14/2018

Executive Summary

Problem Statement:
Understand what drives productivity & identify ways to improve truck productivity

Objective:
Build a model to predict productivity (average sales per month) and use it to recommend next steps

Two-Part Analytical Methodology:
- Feature Creation
- Predictive Modeling

Outcomes/Insights:
Identified key predictors of productivity and relationships between those predictors and productivity

Data Overview and Feature Engineering

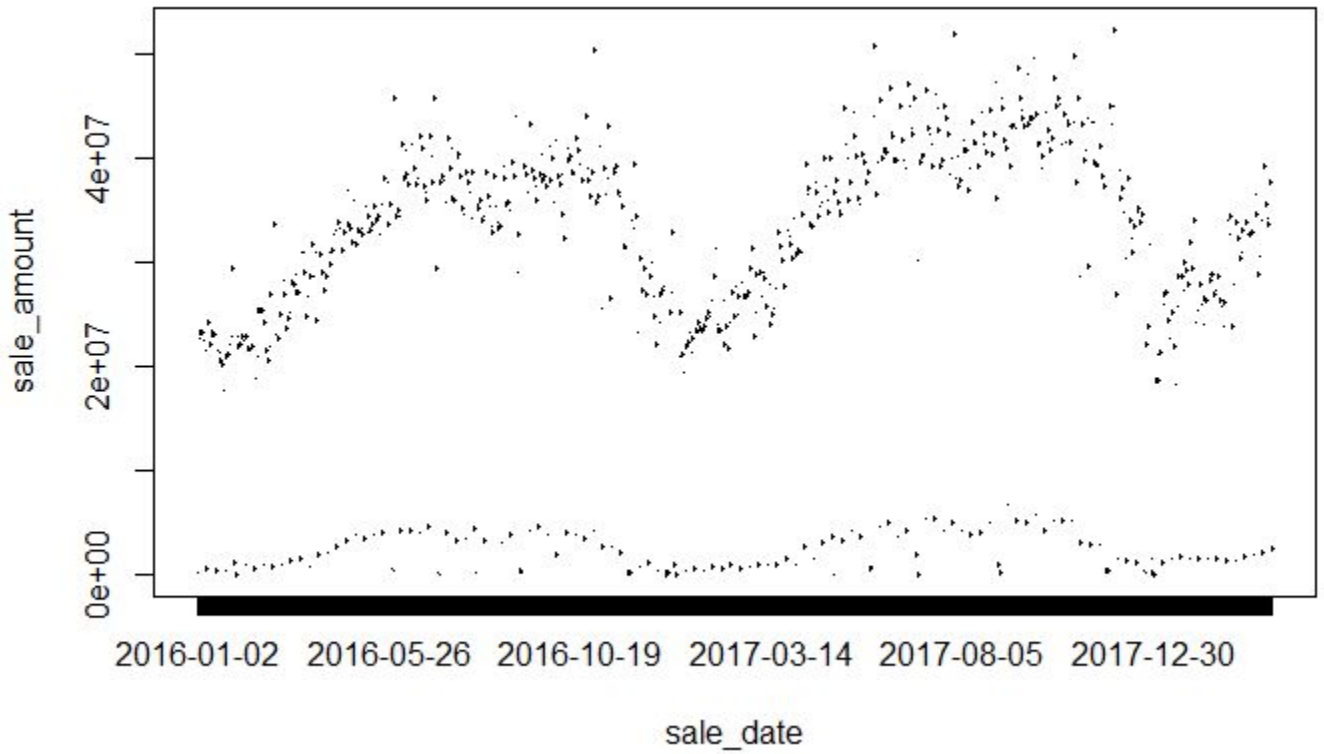
Data set: details of approximately 7,000 trucks in fleet table
Data aggregated for each truck by tying together the route_stop, route, and sales tables

Numeric	Categorical	Spatial/Temporal
Allocation charge	Region	Latitude & Longitude
Insurance charge	District	Time depart & arrive
Repairs	Branch	Model year
Current meter	Unit type	Sales date
Empty/Gross weights	Replace	

Features created:

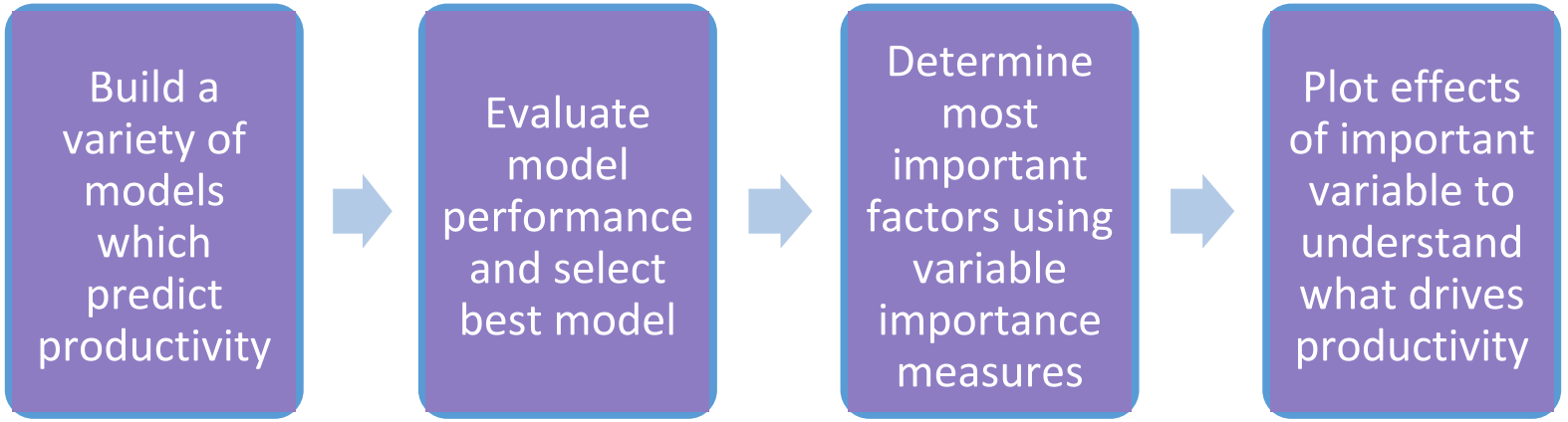
Numeric	Spatial/Temporal
Average routes per day	Average distance traveled
% of morning departures	Average time taken
% of winter season departures	
% of weekday departures	

Total sale amount by date



Modeling Overview

General modeling approach:



Models evaluated:

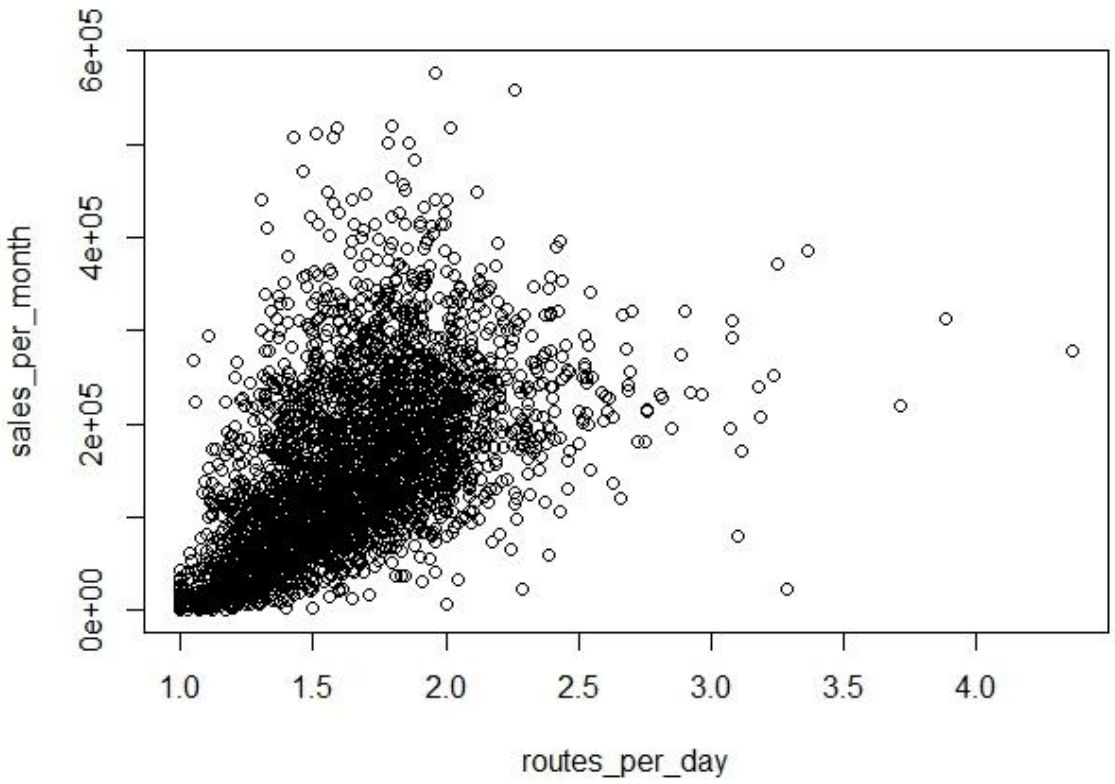
Model	Pros	Cons
Linear Regression	Fast, easily interpretable	Lower accuracy for non-linear data
Random Forest	High accuracy, interpretability, speed	

Linear Regression Model Evaluation

Linear Regression identified several important variables:

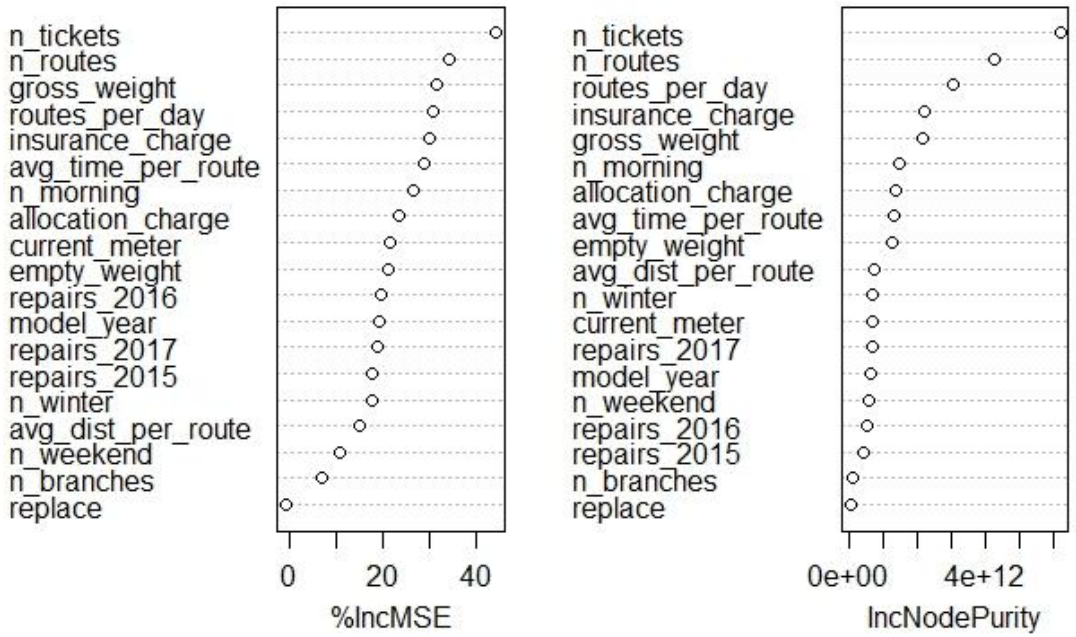
- Average time per route
- Average distance per route
- **Average routes per day**
- Allocation charge
- Unit type (Crane)
- **Repairs**
- Winter deliveries
- Morning deliveries
- **Number of tickets**
- Number of routes
- Number of branches

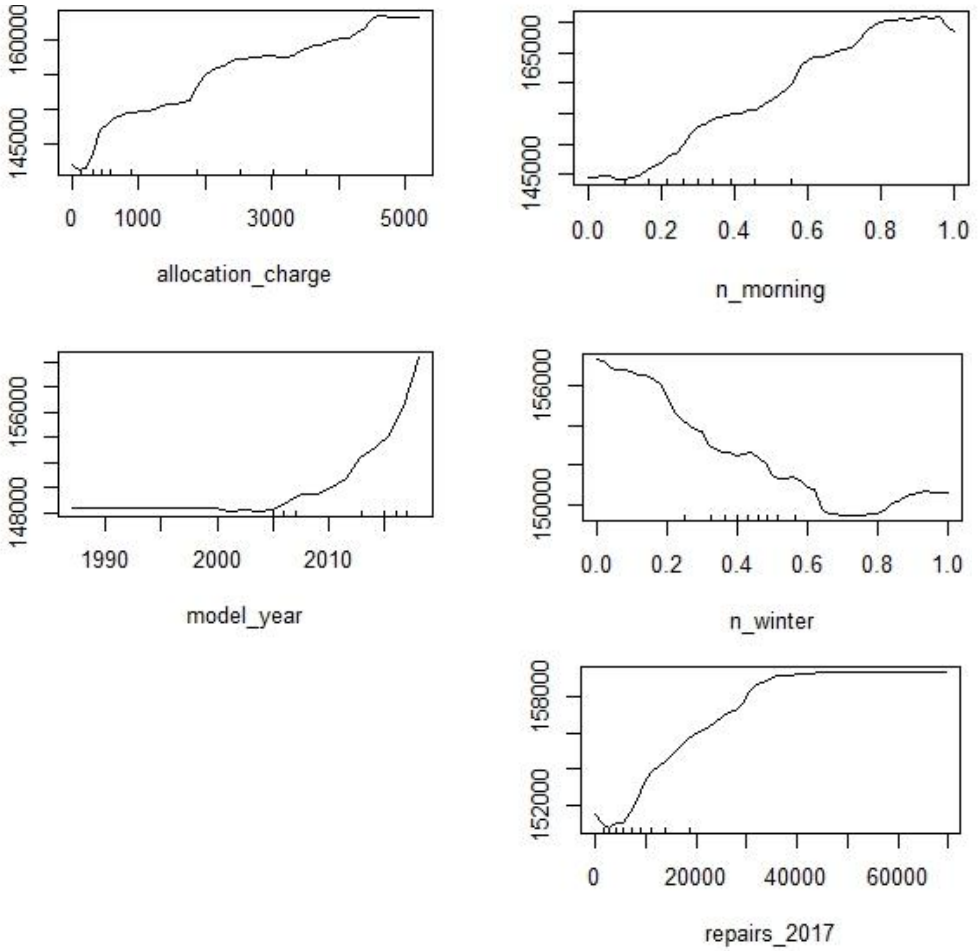
Linear regression R^2: 0.74



Random Forest Model Evaluation

Random Forest identified several important variables:





Random Forest R^2: 0.79

Partial dependency plots reveal how these values increase and/or decrease the sales per month

Recommendations

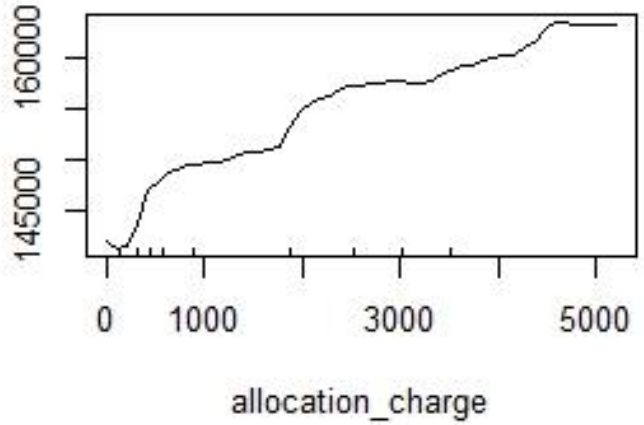
In order to increase productivity:

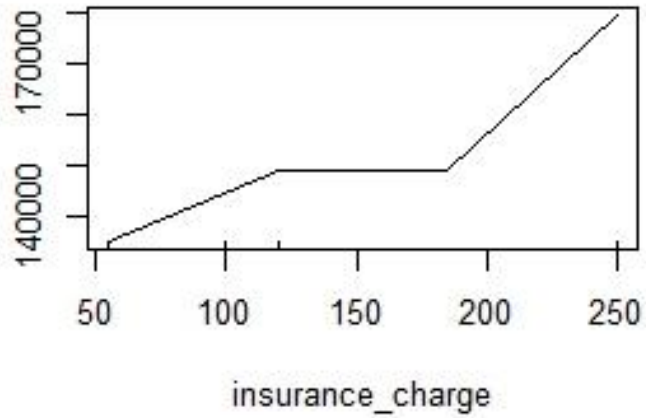
- Increase distance traveled
- Increase time taken
- Repair trucks as soon as possible when necessary
- Schedule more trucks to depart early in the morning
- Schedule fewer trucks during winter season
- Schedule fewer trucks during weekends
- Use trucks that are newer models

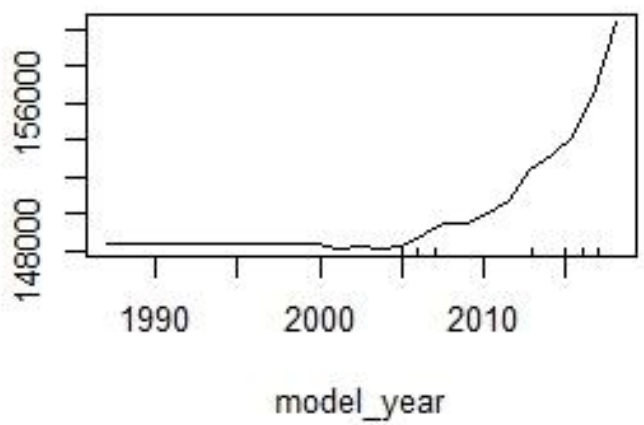
Appendix

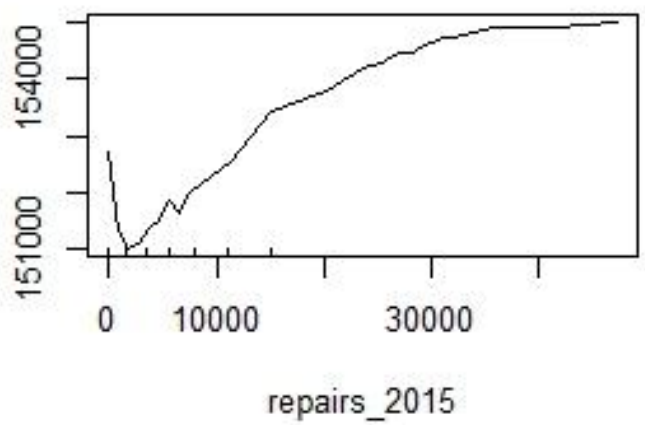
Linear Regression

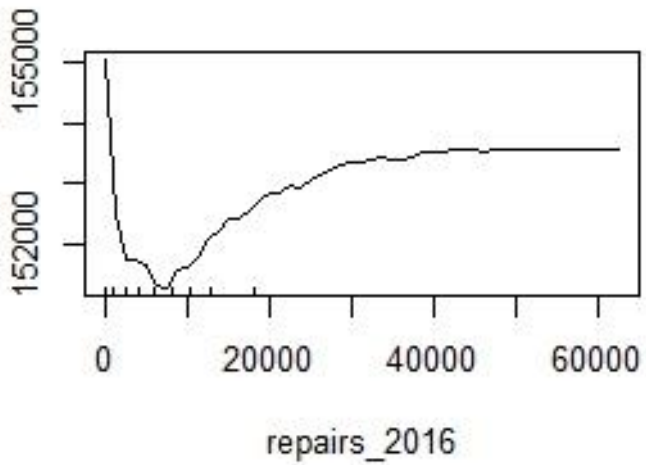
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-4.868e+05	7.290e+05	-0.668	0.50437	
unit_typeConveyor	3.734e+03	5.092e+03	0.733	0.46342	
unit_typeCrane	7.067e+04	8.765e+03	8.062	1.05e-15	***
unit_typeDrywall Boom	1.433e+04	5.468e+03	2.620	0.00883	**
unit_typeFlatbed SA	2.897e+03	5.040e+03	0.575	0.56550	
unit_typeFlatbed TA	2.128e+04	5.002e+03	4.255	2.15e-05	***
unit_typeKnuckle Boom	1.709e+04	5.282e+03	3.236	0.00123	**
unit_typePick-up	1.565e+02	7.017e+03	0.022	0.98221	
unit_typeRear Boom	2.925e+03	7.717e+03	0.379	0.70466	
unit_typeSemi Tractor	5.276e+04	2.361e+04	2.235	0.02547	*
allocation_charge	1.123e+01	1.448e+00	7.754	1.20e-14	***
insurance_charge	2.439e+02	1.567e+02	1.556	0.11973	
model_year	1.927e+02	3.628e+02	0.531	0.59531	
repairs_2015	2.977e-01	1.890e-01	1.575	0.11534	
repairs_2016	-1.741e-01	1.642e-01	-1.060	0.28901	
repairs_2017	7.441e-01	1.554e-01	4.789	1.75e-06	***
current_meter	1.842e-03	3.518e-03	0.524	0.60056	
replace1	-5.267e+02	3.706e+03	-0.142	0.88700	
n_routes	9.503e+01	1.470e+01	6.466	1.16e-10	***
n_branches	-5.765e+03	2.091e+03	-2.758	0.00585	**
n_tickets	9.275e+01	4.925e+00	18.830	< 2e-16	***
n_morning	2.093e+04	7.011e+03	2.985	0.00285	**
n_winter	-2.690e+04	6.380e+03	-4.217	2.55e-05	***
n_weekend	-4.679e+04	5.007e+04	-0.935	0.35005	
avg_dist_per_route	8.557e+00	1.728e+00	4.951	7.79e-07	***
avg_time_per_route	1.222e+02	1.880e+01	6.501	9.27e-11	***
routes_per_day	4.065e+04	4.474e+03	9.087	< 2e-16	***

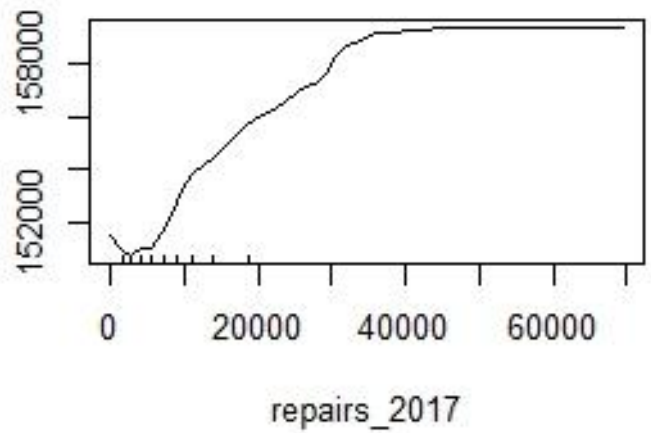

145000
160000
0
1000
3000
5000
allocation_charge

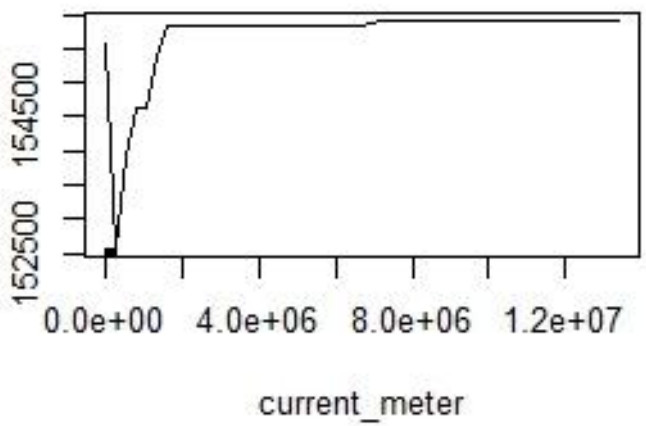

140000
170000
50
100
150
200
250
insurance_charge

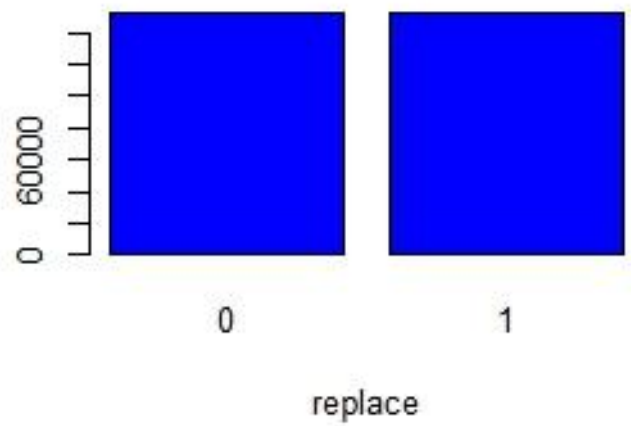

148000
156000
1990
2000
2010
model_year

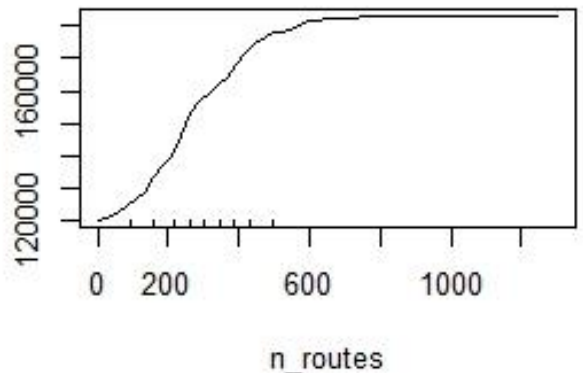

151000
154000
0
10000
30000
repairs_2015

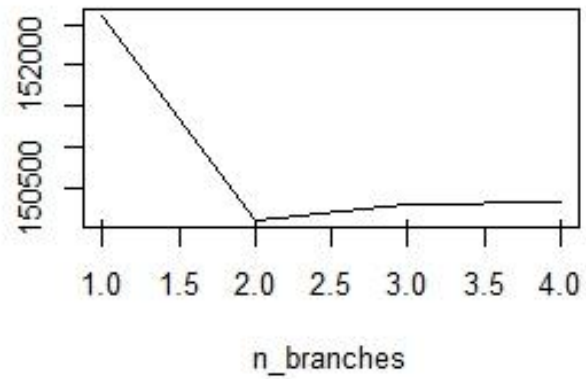

155000
152000
0
20000
40000
60000
repairs_2016

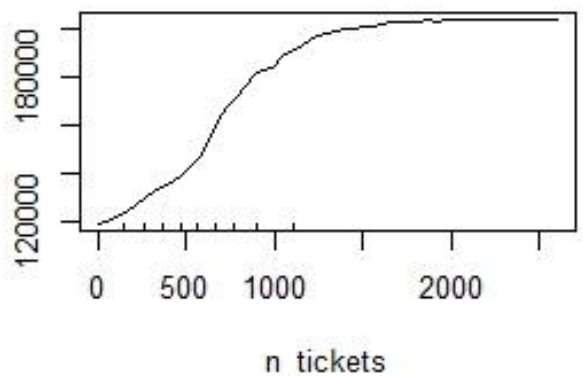

158000
152000
0
20000
40000
60000
repairs_2017

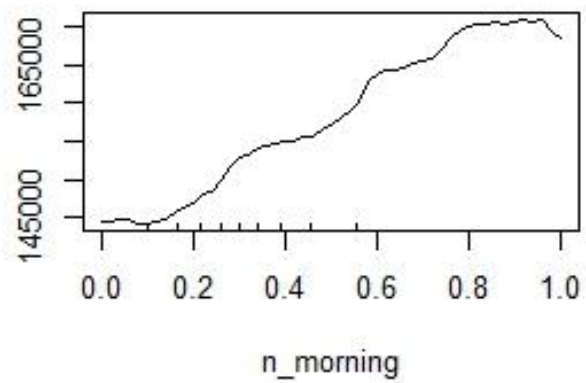

154500
152500
0.0e+00
4.0e+06
8.0e+06
1.2e+07
current_meter

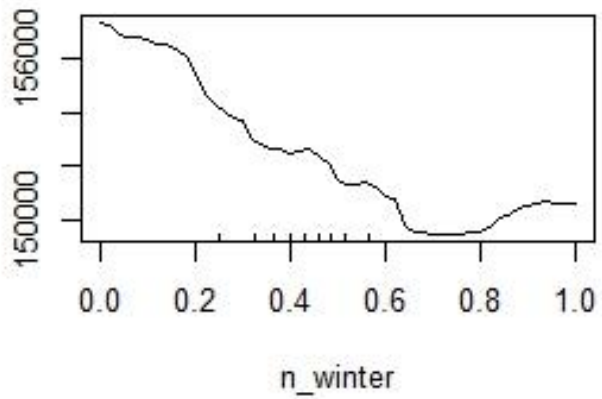

60000
0
0
1
replace

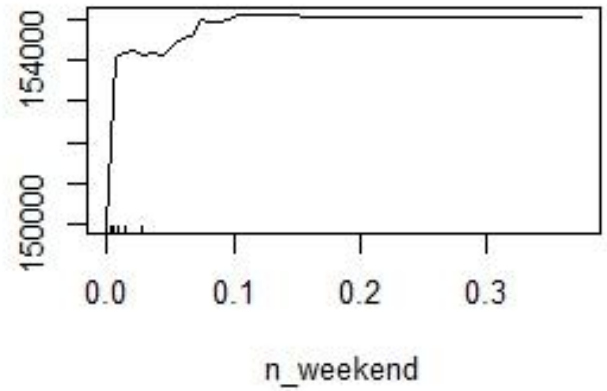

120000
160000
0
200
600
1000
n_routes

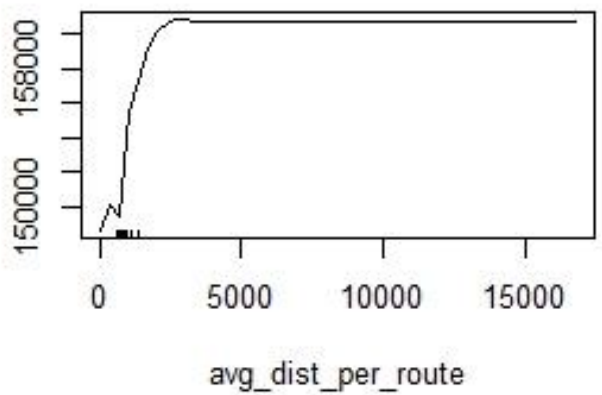

150500
152000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
n_branches

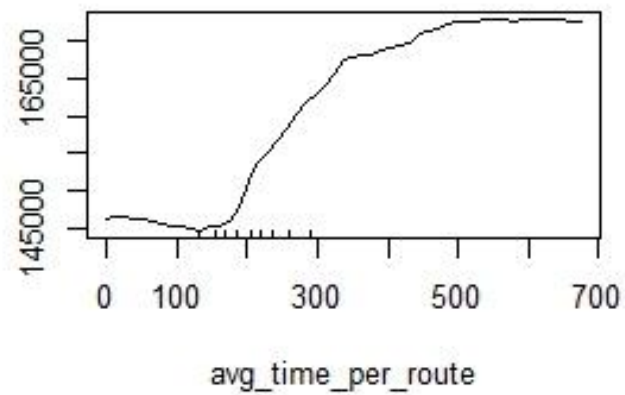

120000
180000
0
500
1000
2000
n_tickets

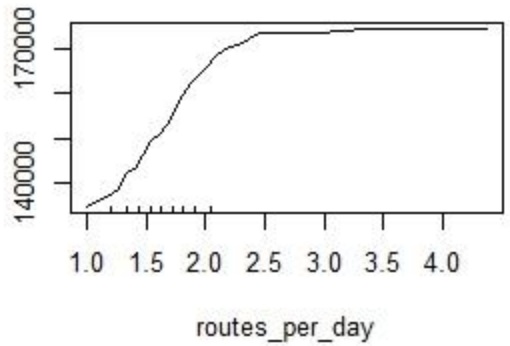

145000
165000
0.0
0.2
0.4
0.6
0.8
1.0
n_morning

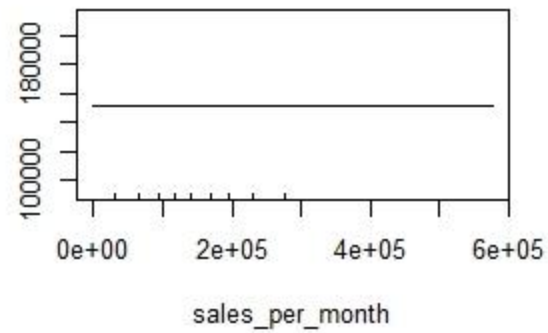

150000
156000
0.0
0.2
0.4
0.6
0.8
1.0
n_winter

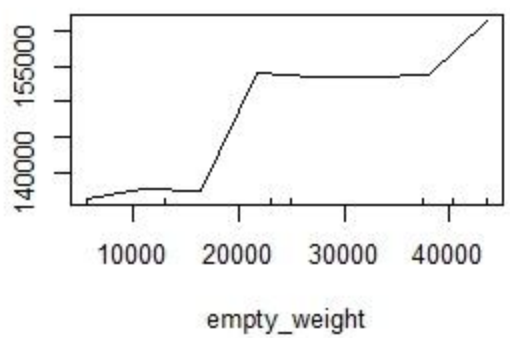

150000
154000
0.0
0.1
0.2
0.3
n_weekend

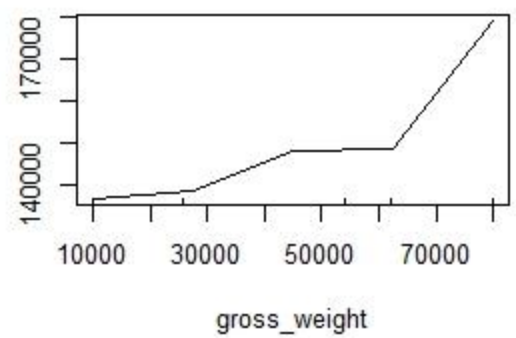

150000
158000
0
5000
10000
15000
avg_dist_per_route


145000
165000
0
100
300
500
700
avg_time_per_route


140000
170000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
routes_per_day


100000
180000
0e+00
2e+05
4e+05
6e+05
sales_per_month


140000
155000
10000
20000
30000
40000
empty_weight


140000
170000
10000
30000
50000
70000
gross_weight