Customer Propensity for Service (Ashok Leyland Ltd (AL))

Debasish Patra (71620022)

Rama Chandra Vattigunta (71620097)

Soubhagya Ranjan Rout (71620070)

Vaibhav Pralhad Mhaske (71620083)

(ISB - CBA Batch 7)

Project Sponsor - J Balasubramanian

Agenda

- Introduction
- Key Points
- Conclusions
- Recommendations

Customer Propensity for Service

- If you order pizza from Pizza Hut today then what is the probability that you will order pizza from Pizza hut next week or next month?

- On similar lines ….

- If a customer comes to service his vehicle to service outlet today, what is the probability that the customer will visit the service outlet next month, or maybe after 3 months ?

- What is **Customer Propensity for Service** ????

Key Points

- Business Problem
- Descriptive and Exploratory Analysis of Data
- Categorization of Visits and Analytical Techniques used
- Comparison of Models
- Results and Interpretation

Business Problem

Problem

- To obtain Propensity / Probability scores against each customer who will visit dealership workshop for his vehicle servicing needs during the defined period of time

Sub Problem

- Prediction of customers to be met by the Service Marketing Executive, Customer’s past service patterns and Services/potential for these Customer vehicles

Data Cleaning and Data Exploration

- Vehicle cost zero or negative
- Service amount is zero or negative
- Visit data had two formats (ddmmyyyy and mmddyyyy). Hence resolved the format issue
- Found common component ids from Sales data and Service data and filtered both the datasets for common component ids
- Merged multiple line items of a service into a single line item of service
- Merged sales data and service data to obtain a single dataset

Feature extraction

	compno	segment	model	lob	docid	kmread	amount		repairtype	region
1	ABE025007Y	Segment 2	Model 90	LOB 4	12070864	22881	1145.22	REPAIR TYPE 13	REPAIR TYPE 22	Region 13
2	ABE025007Y	Segment 2	Model 90	LOB 4	12385555	26897	11646.92	REPAIR TYPE 13	REPAIR TYPE 22	Region 13
3	ABE025007Y	Segment 2	Model 90	LOB 4	14048956	47668	12432.42		REPAIR TYPE 13	Region 13
4	ABE025007Y	Segment 2	Model 90	LOB 4	15073898	48000	22592.50		REPAIR TYPE 13	Region 13
5	ABE025007Y	Segment 2	Model 90	LOB 4	15120319	48045	2191.40		REPAIR TYPE 13	Region 13
6	ABE025007Y	Segment 2	Model 90	LOB 4	15129685	48046	240.07		REPAIR TYPE 13	Region 13

	vehiclecost	billdate	dateofsale	vstdate	daysino	datediff	serviceno	group	dtdflag	servdiff
1	1400000	2015-10-27	2014-12-13	2015-10-27	0	318	1	2014 Dec	0	318
2	1400000	2015-12-15	2014-12-13	2015-12-14	1	366	2	2014 Dec	318	48
3	1400000	2016-07-18	2014-12-13	2016-07-18	0	583	3	2014 Dec	366	217
4	1400000	2016-12-03	2014-12-13	2016-11-28	5	716	4	2014 Dec	583	133
5	1400000	2016-12-04	2014-12-13	2016-12-04	0	722	5	2014 Dec	716	6
6	1400000	2016-12-08	2014-12-13	2016-12-05	3	723	6	2014 Dec	722	1

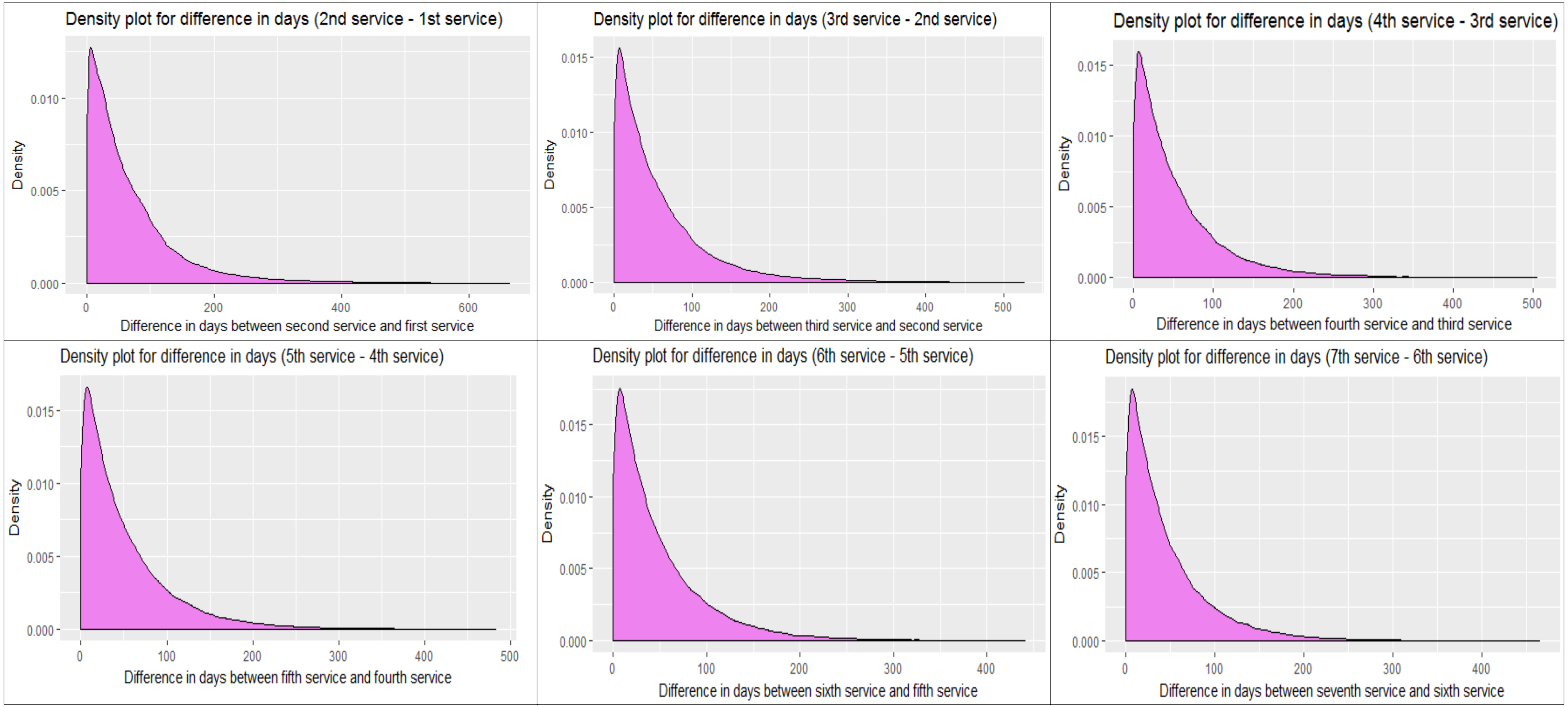
Difference in days between consecutive services

Bill date – Visit date

Number in days between service and Date of Sale

Service no of services for vehicles

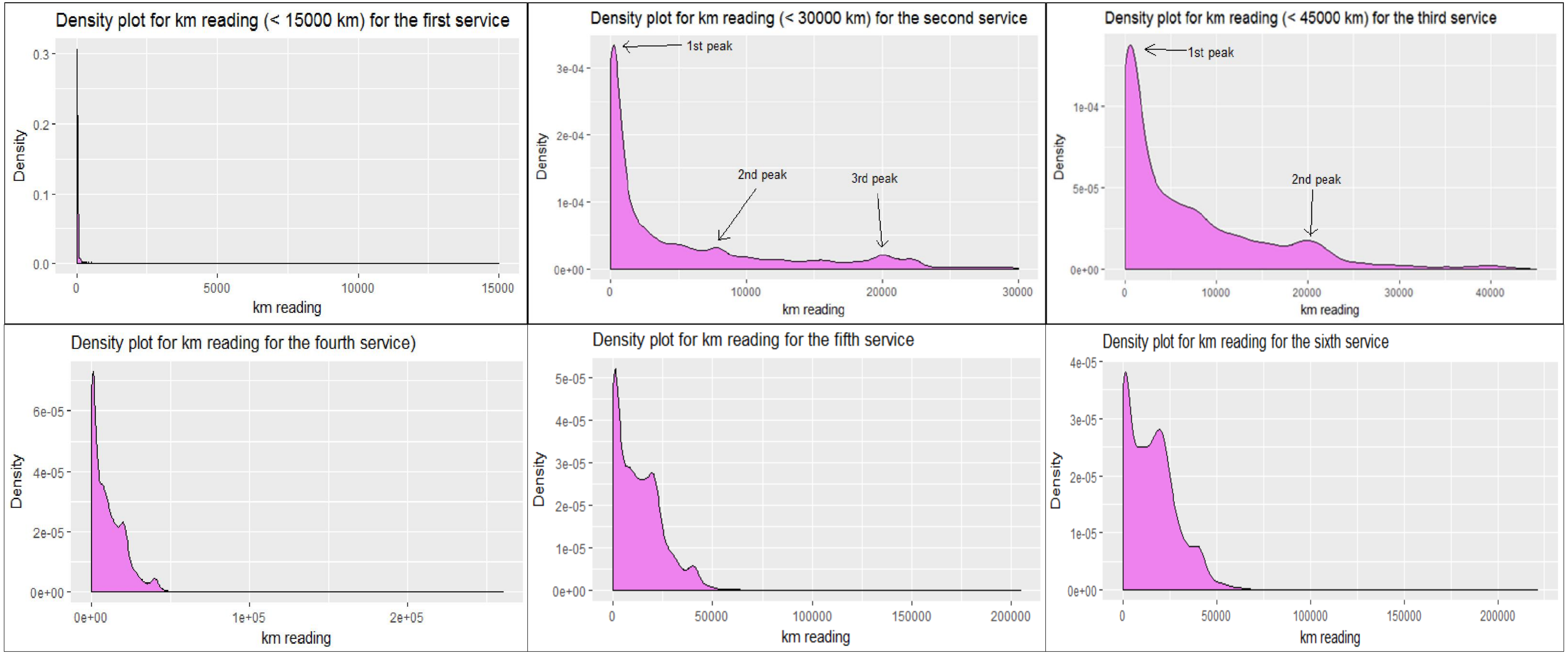
Density plot for difference in days between consecutive services from 2nd service to 7th service



Summary of Difference in days of consecutive services

Difference in days of consecutive services	Median (Difference in days)
2nd service – 1st service	45
3rd service – 2nd service	38
4th service – 3rd service	36
5th service – 4th service	35
6th service – 5th service	33
7th service – 6th service	31

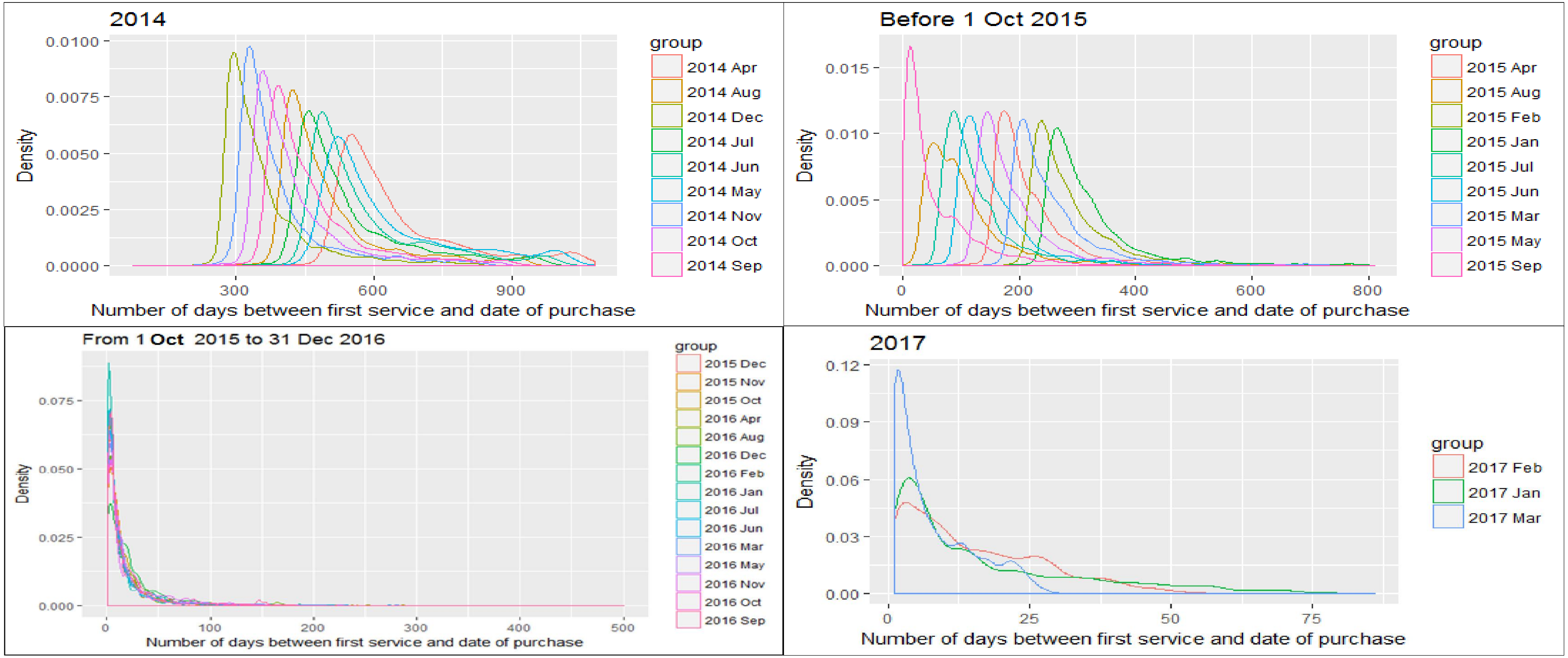
Density plot for km reading of 1st to 6th service (Date of sale from 1st Oct 2015)



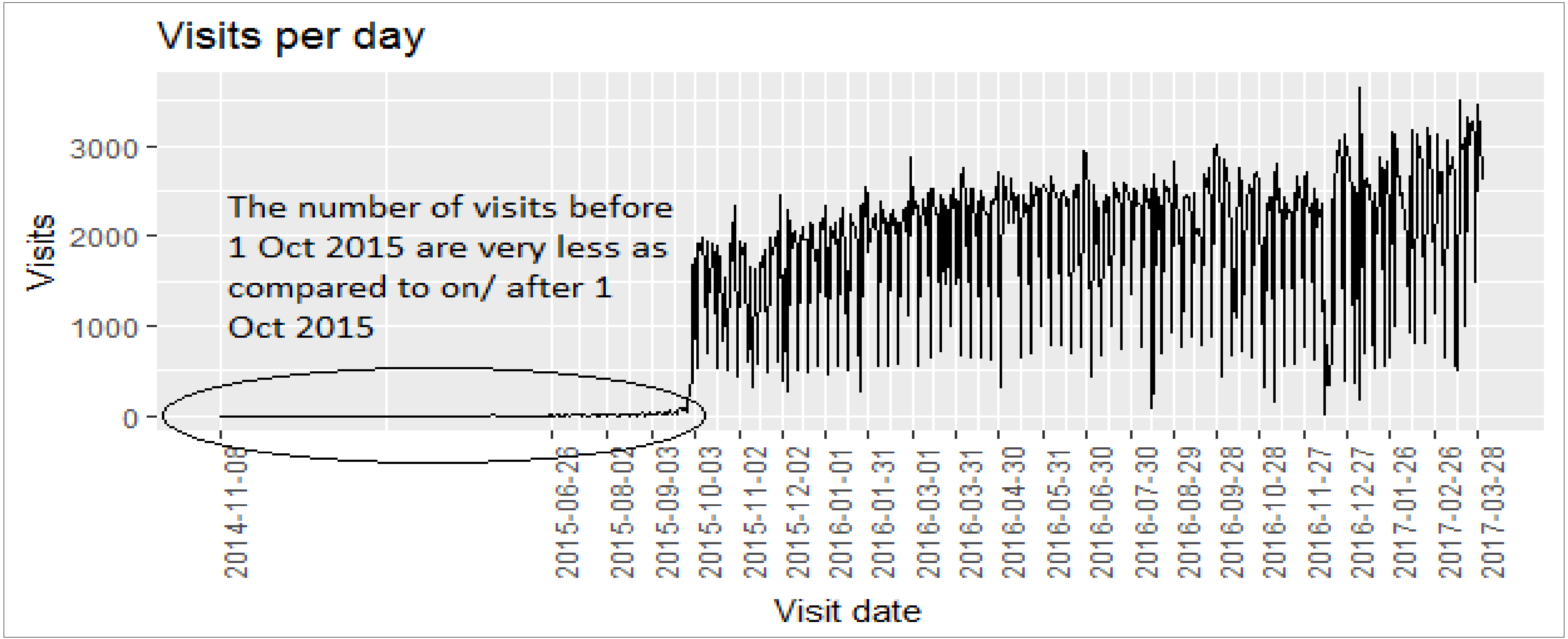
Summary for km reading of 1st to 6th service (Date of sale from 1st Oct 2015)

Service number	Median (km reading)
1	10
2	2317
3	9468
4	23456
5	33202
6	40158

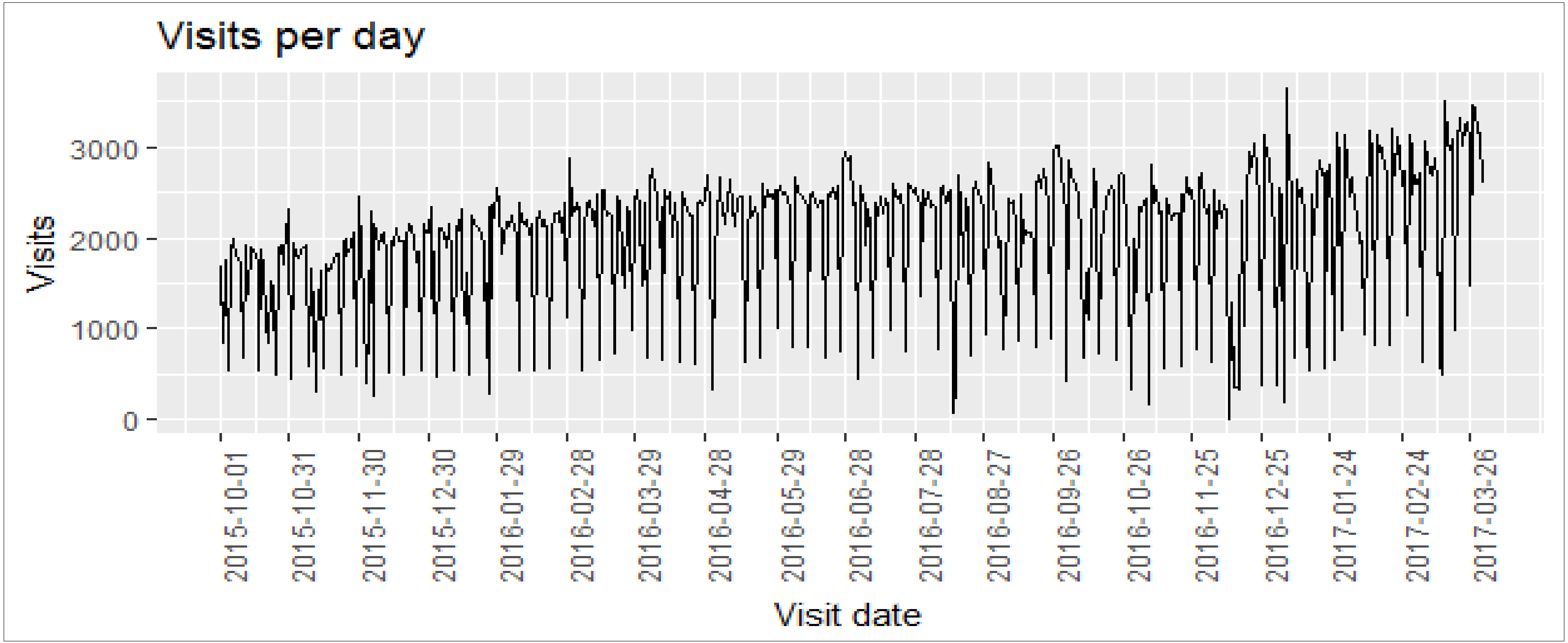
Missing Service data (Density plot for first service (date of sale month-wise))



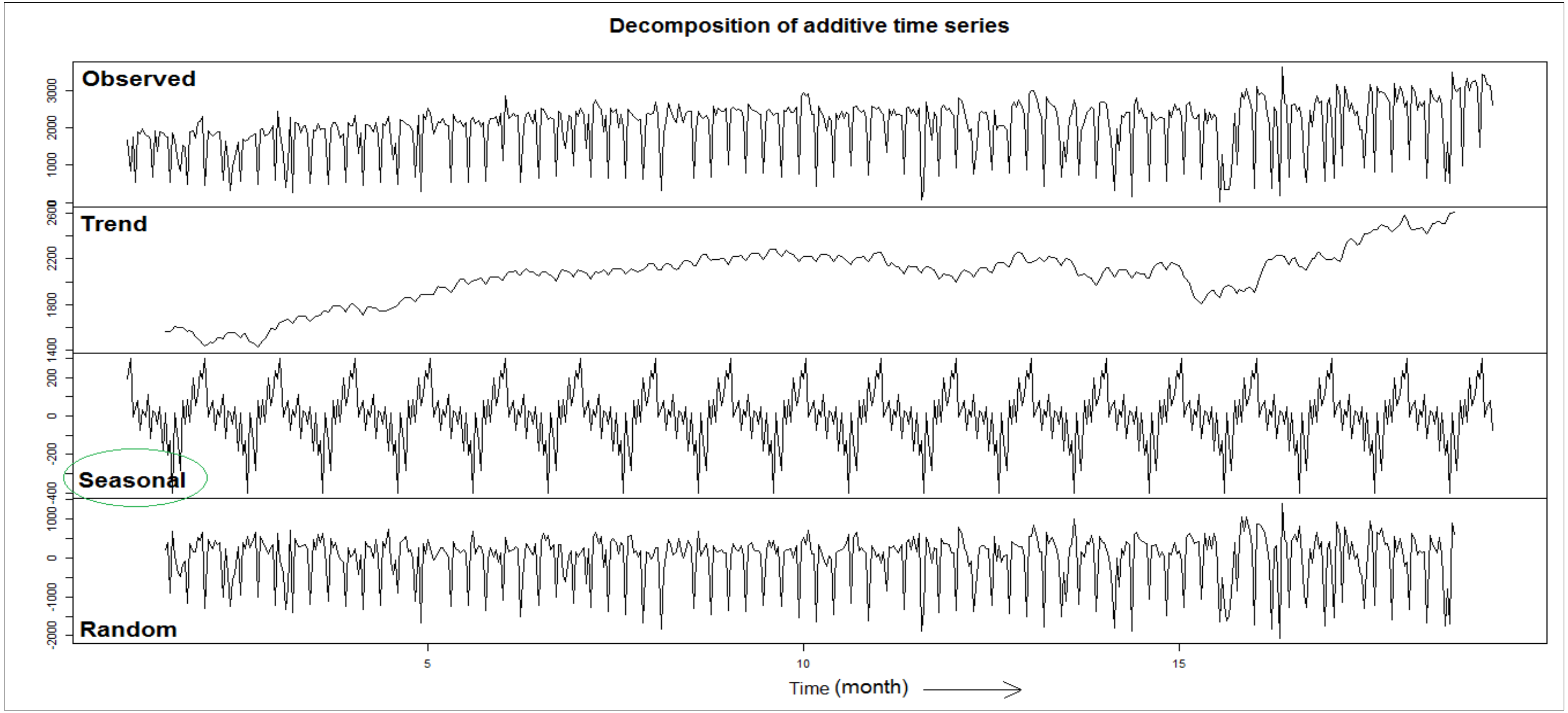
Seasonality in Visits



Seasonality in Visits contd’



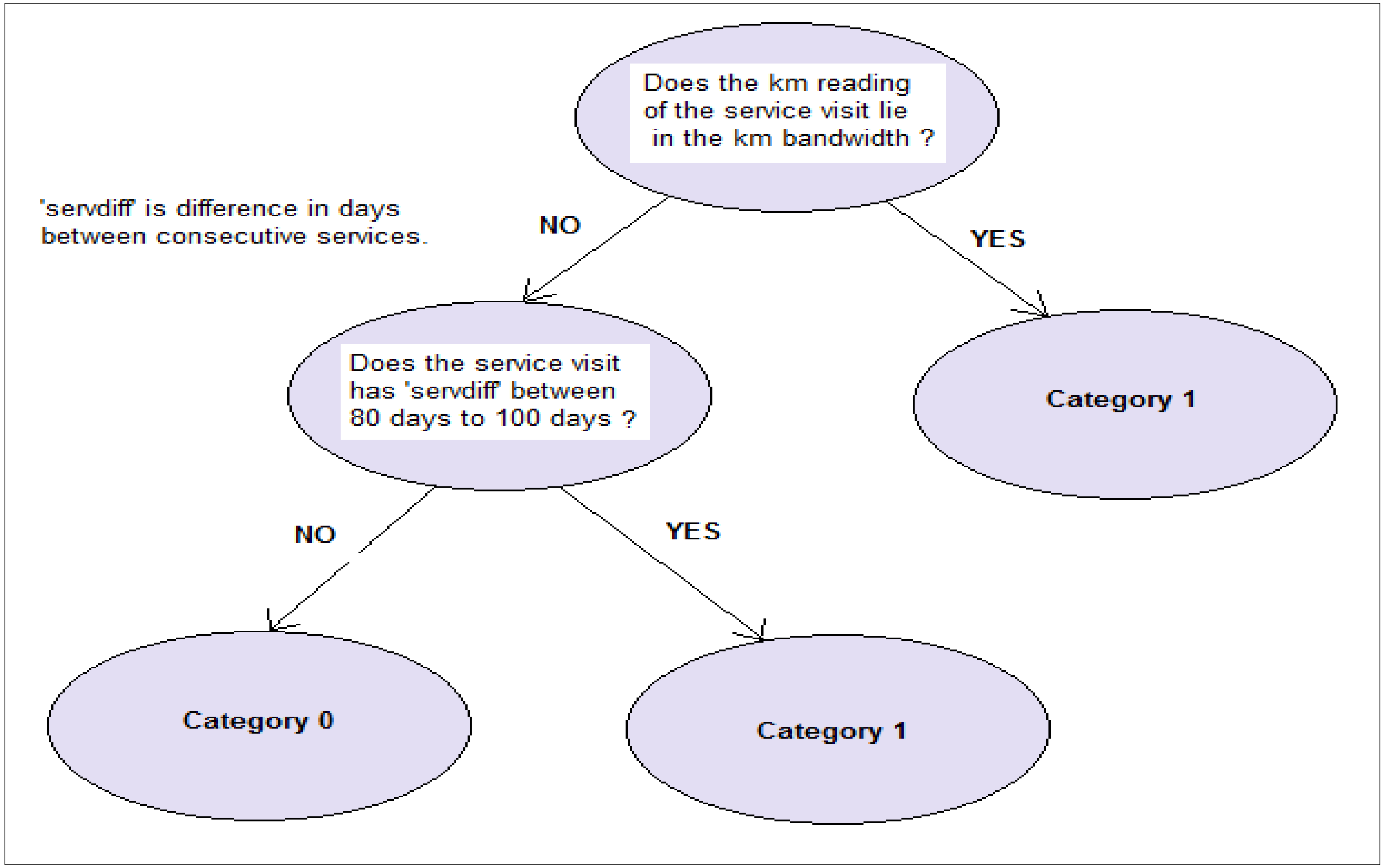
Seasonality in Visits



More features added

- Create categories for variable model
- Added lag variables of service data to each observation
- Created categories for variable lag repair type into individual lag repair types taking values ‘1’ or ‘0’
- Time index tmonth
  - t = 10 for Oct 2015
- Sine seasonality = sin((2*∏*t)/12)
- Cosine seasonality = cos((2*∏*t)/12)
- Dependent Variable is Visit
  - Can takes values ‘1’ or ‘0’

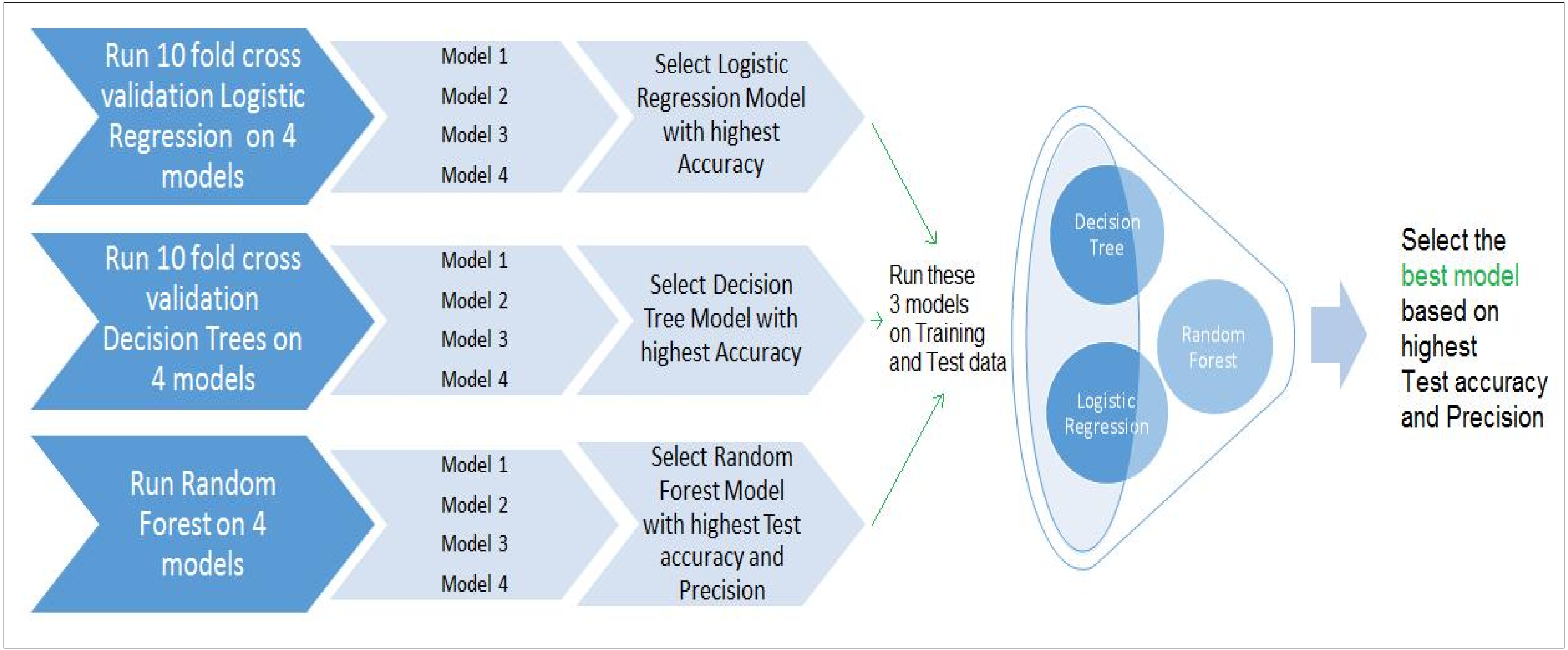
Categorization of Visits (Method 1)



Formula for Models (Method 1)

- **Model 1**
  - visit~segment+lob+region+model571+model561+model562+lagkmread+lagamount+lgrtpaid+lgrtamc+lgrtancillary+lgrtfoc+lgrtfreeservice+lgrtwarrant+lagdatediff+lagdaysino+lagservdiff+tmonth
- **Model 2**
  - visit~segment+lob+region+model571+model561+model562+lagkmread+lagamount+lgrtpaid+lgrtamc+lgrtancillary+lgrtfoc+lgrtfreeservice+lgrtwarrant+lagdatediff+lagdaysino+lagservdiff+tmonth+tmonthsins+tmonthcoss
- **Model 3**
  - visit~segment+lob+region+model571+model561+model562+log(lagkmread+1)+log(lagamount+1)+lgrtpaid+lgrtamc+lgrtancillary+lgrtfoc+lgrtfreeservice+lgrtwarrant+lagdatediff+lagdaysino+lagservdiff+tmonth
- **Model 4**
  - visit~segment+lob+region+model571+model561+model562+log(lagkmread+1)+log(lagamount+1)+lgrtpaid+lgrtamc+lgrtancillary+lgrtfoc+lgrtfreeservice+lgrtwarrant+lagdatediff+lagdaysino+lagservdiff+tmonth+tmonthsins+tmonthcoss

Predictive Modelling (Method 1)



Cross Validation (Method 1)

Method	Model	Accuracy (obtained from cross validation)
10 fold cross validation (Logistic Regression)	Model 1	0.7097402
10 fold cross validation (Logistic Regression)	Model 2	0.7082656
10 fold cross validation (Logistic Regression)	Model 3	0.7141222
10 fold cross validation (Logistic Regression)	Model 4	0.7136789

Method	Model	Accuracy (obtained from cross validation)
10 fold cross validation (Decision Trees)	Model 1	0.7450177
10 fold cross validation (Decision Trees)	Model 2	0.7452369
10 fold cross validation (Decision Trees)	Model 3	0.7450324
10 fold cross validation (Decision Trees)	Model 4	0.7444610

Random Forest Models (Method 1)

Method	Model	Train Accuracy	Test Accuracy	Precision
Random Forest	Model 1	96.86592	78.23698	70.82284
Random Forest	Model 2	97.35192	78.17356	70.82974
Random Forest	Model 3	96.86153	78.30385	70.91710
Random Forest	Model 4	97.51001	78.15977	70.57703

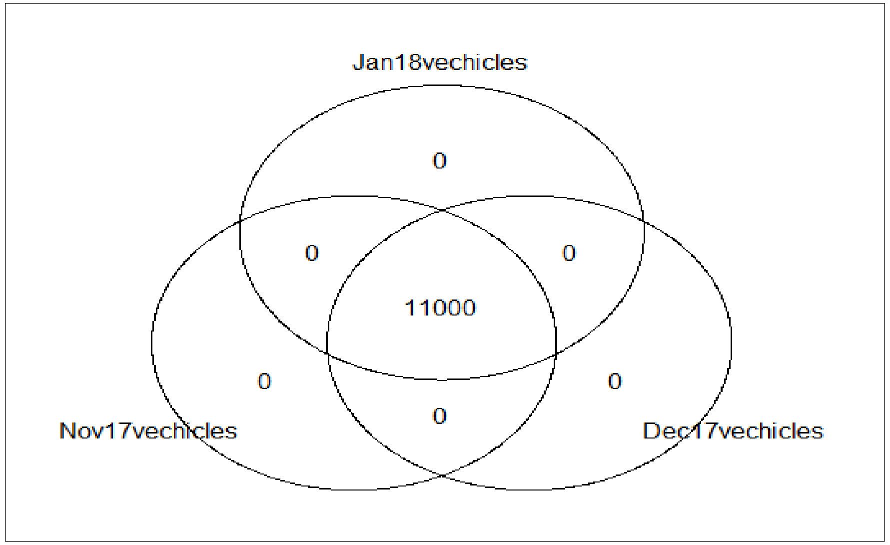
Comparison of Models (Method 1)

Method	Model	Train Accuracy	Test Accuracy	Precision
Logistic Regression	Model 3	71.41309	71.60756	60.58116
Decision Trees	Model 2	75.94352	76.12093	65.05978
Random Forest	Model 3	96.86153	78.30385	70.91710

loglagkmread most important variable (MeanDecreaseGini)

Prediction of Customers for Months Nov 17, Dec 17 and Jan 18

- We predict customers to visit outlets for months Nov 17, Dec 17 and Jan 18
- Take top 11000 vehicles from 130993 vehicles based on probability scores for months Nov 17, Dec 17 and Jan 18

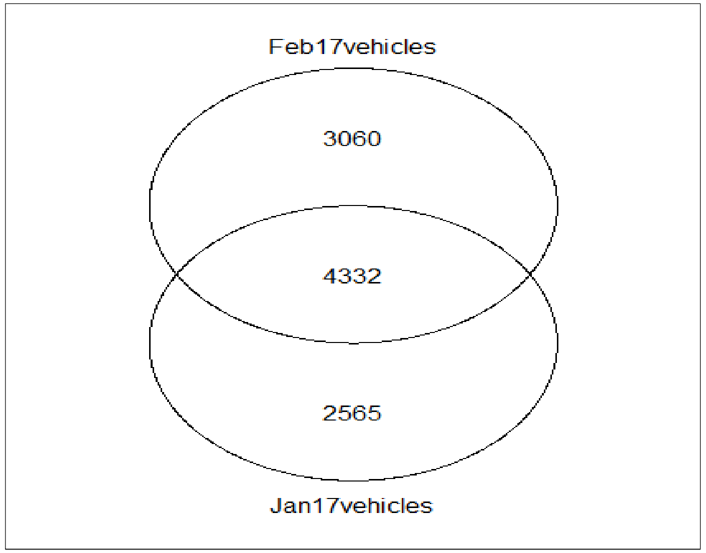


- Hence, we need to recalibrate the model

Recalibration of Model

Method	Model	Data	Train accuracy	Test accuracy	Precision
Random Forest	Model 3	Till Dec 16	97.01166	78.82469	68.25576
Random Forest	Model 3	Till Jan 17	96.86487	78.81995	68.41591

- We predict vehicles to visit outlets in next month
  - Model having data till Dec 16, we predict for Jan 17
  - Model having data till Jan 17 , we predict for Feb 17

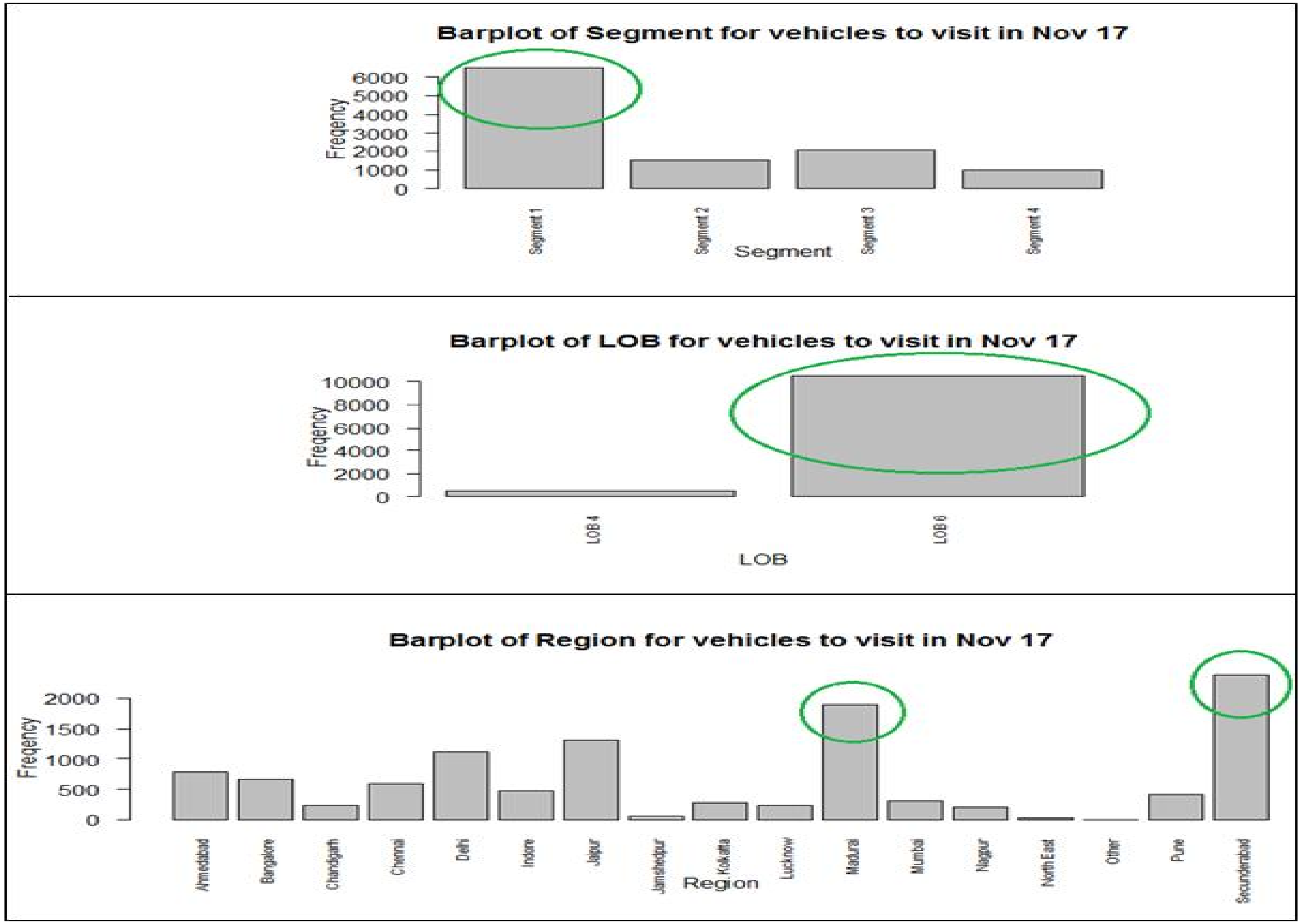


Top 10 vehicles predicted for month of Nov 17

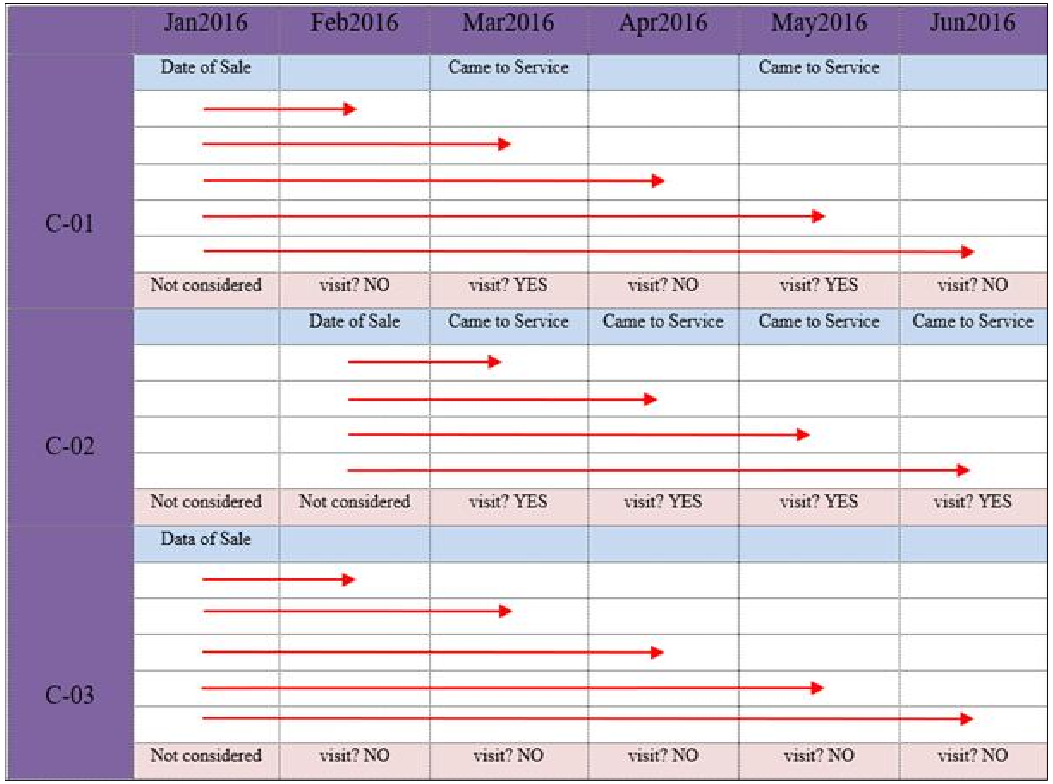
Component No	Probability Score
FBEZ408375	1
FBEZ408507	1
FBEZ409747	1
FBPZ111519	1
FCPZ107967	1
FDEZ201595	1
FDEZ403504	1
FDEZ404741	1
FDEZ404748	1
FDEZ405243	1

```
> dtmdl[which(dtmdl$compno == "FBEZ408507" ),c(1,4,5,6,10,11:13,15,16)]
         compno kmread   amount repairtype    vstdate daysino datediff serviceno servdiff tmonth
3878 FBEZ408507   1072  8176.95       PAID 2016-05-18       1       95         1       95     17
3879 FBEZ408507   1198  2854.75       PAID 2016-05-19       0       96         2        1     17
3880 FBEZ408507   1582 24393.68       PAID 2016-08-23       0      192         3       96     20
3881 FBEZ408507   1590  6353.70       PAID 2017-01-30       0      352         4      160     25
3882 FBEZ408507   2521 28843.66       PAID 2017-03-07       0      388         5       36     27
```

Additional Insights (Method 1)



Categorization of Visits (Method 2)



Month	Feb-16	Mar-16	Apr-16	May-16	Jun-16
Visit	No	Yes	No	Yes	No

Derived Variables (Method 2)

- KM difference between last two services
- Maximum amount spent in any month since date of sale
- Average amount spent in any month when vehicle came for the service
- Average day of stay in outlet per month, when vehicle came for service
- Maximum Bill date - Minimum Visit date, this variable will be different from stays in outlet, if vehicle came for service multiple time
- Age of the vehicle in months
- Month duration till the vehicle came for last service
- Number of months when the vehicle came for service

Formula for Models (Method 2)

- **Model 1**
  - Derived variables are used as independent variables.
  - is.visit~KM.reading+km.diff+no.of.ser.mon+service.cnt+amount+avg.amt.spent+max.amt.spent+VEHICLE.COST+mean.out.stay+mn.outlet.stay+avg.outlet.stay+vehicle.age.mon+last.service.age.mon+is.paid+is.warranty+is.foc+is.free.service+is.presales+is.amc+is.ancillary+LINE.OF.BUSINESS+SEGMENT+u.model+new.region
- **Model 2**
  - log variables for "KM.reading", "km.diff", "amount", "avg.amt.spent" and "max.amt.spent" added to the Model 1.
  - is.visit~KM.reading+log.KM.reading+km.diff+log.KM.diff+no.of.ser.mon+service.cnt+amount+log.amount+avg.amt.spent+log.avg.amt.spent+max.amt.spent+log.max.amt.spent+VEHICLE.COST+mean.out.stay+mn.outlet.stay+avg.outlet.stay+vehicle.age.mon+last.service.age.mon+is.paid+is.warranty+is.foc+is.free.service+is.presales+is.amc+is.ancillary+LINE.OF.BUSINESS+SEGMENT+u.model+new.region
- **Model 3**
  - Using AIC criterion for obtaining the best subset from Model 2. Dropping "is.ancillary", "LINE.OF.BUSINESS", " VEHICLE.COST " and "KM.reading" as obtained from the analysis.
  - is.visit~log.KM.reading+km.diff+log.KM.diff+no.of.ser.mon+service.cnt+amount+log.amount+avg.amt.spent+log.avg.amt.spent+max.amt.spent+log.max.amt.spent+mean.out.stay+mn.outlet.stay+avg.outlet.stay+vehicle.age.mon+last.service.age.mon+is.paid+is.warranty+is.foc+is.free.service+is.presales+is.amc+SEGMENT+u.model+new.region
- **Model 4**
  - Remove ingnificant variables "no.of.ser.mon" and "log.max.amt.spent" from Model 3.
  - is.visit~log.KM.reading+km.diff+log.KM.diff+service.cnt+amount+log.amount+avg.amt.spent+log.avg.amt.spent+max.amt.spent+mean.out.stay+mn.outlet.stay+avg.outlet.stay+vehicle.age.mon+last.service.age.mon+is.paid+is.warranty+is.foc+is.free.service+is.presales+is.amc+SEGMENT+u.model+new.region

Cross Validation (Method 2)

Method	Model	ROC	Sensitivity	Specificity
Logistic Regression (10-fold CV)	Model 1	0.7418075	0.6729161	0.695246
Logistic Regression (10-fold CV)	Model 2	0.8568601	0.8002683	0.7510183
Logistic Regression (10-fold CV)	Model 3	0.8569293	0.8005667	0.7506284
Logistic Regression (10-fold CV)	Model 4	0.8570705	0.8008957	0.7510791

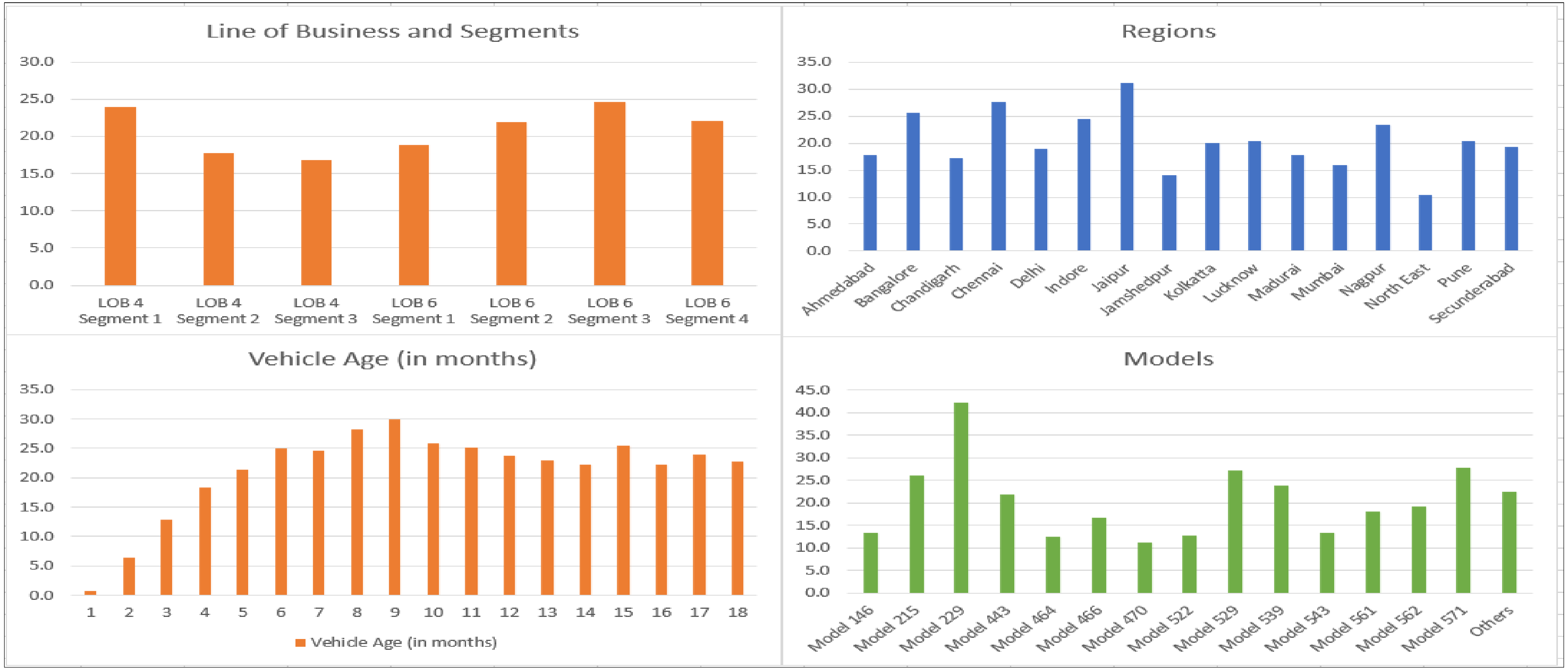
Random Forest Models (Method 2)

Method	Model	ROC	Sensitivity	Specificity
Random Forest	Model 1	0.9541772	0.8622745	0.8855686
Random Forest	Model 2	0.9550518	0.8631373	0.8871373
Random Forest	Model 3	0.9550529	0.8643137	0.8812549
Random Forest	Model 4	0.9558711	0.8680000	0.8803922

Comparison of Models (Method 2)

Method	Model	Test Accuracy	Precision
Logistic Regression	Model 4	0.7771	0.7649
Random Forest	Model 4	0.8724	0.878

Additional Insights (Method 2)



Cost Benefit Analysis

Predicted customers per month	Customers converted (75%)	Average Revenue of service (INR)	Total yearly Revenue of service (INR crore)	Total yearly profit (30%) in INR crore
11000	8250	5968	59.08	17.72

Customer Service Executives	Cost of Customer Service Executives (400000 INR per year)	IT infrastructure cost (INR)	Total Cost (INR crore)	Total Net Profit (INR crore)
110	44000000	10000000	5.4	**12.32**

Conclusions and Recommendations

- Predicted customers can give an impetus to the service business of the company
- Increase service revenue
- Create an innovative business process for customer identification for service business
- Additional Insights - efficiently manage the service marketing executives and also better manage supply chain for spares
- Use recent data of around 18 – 24 months and run the best selection model process every month to predict the customers next month

Future Actions and Acknowledgements

- Future Actions
  - More classifications model can be implemented
  - Improvement in Accuracy

- Acknowledgements
  - Faculty Advisor - Prof. Manish Gupta
  - Project Sponsor - J. Balasubramanian
  - Project Mentor – Santa Suman Dutta

- Thank You