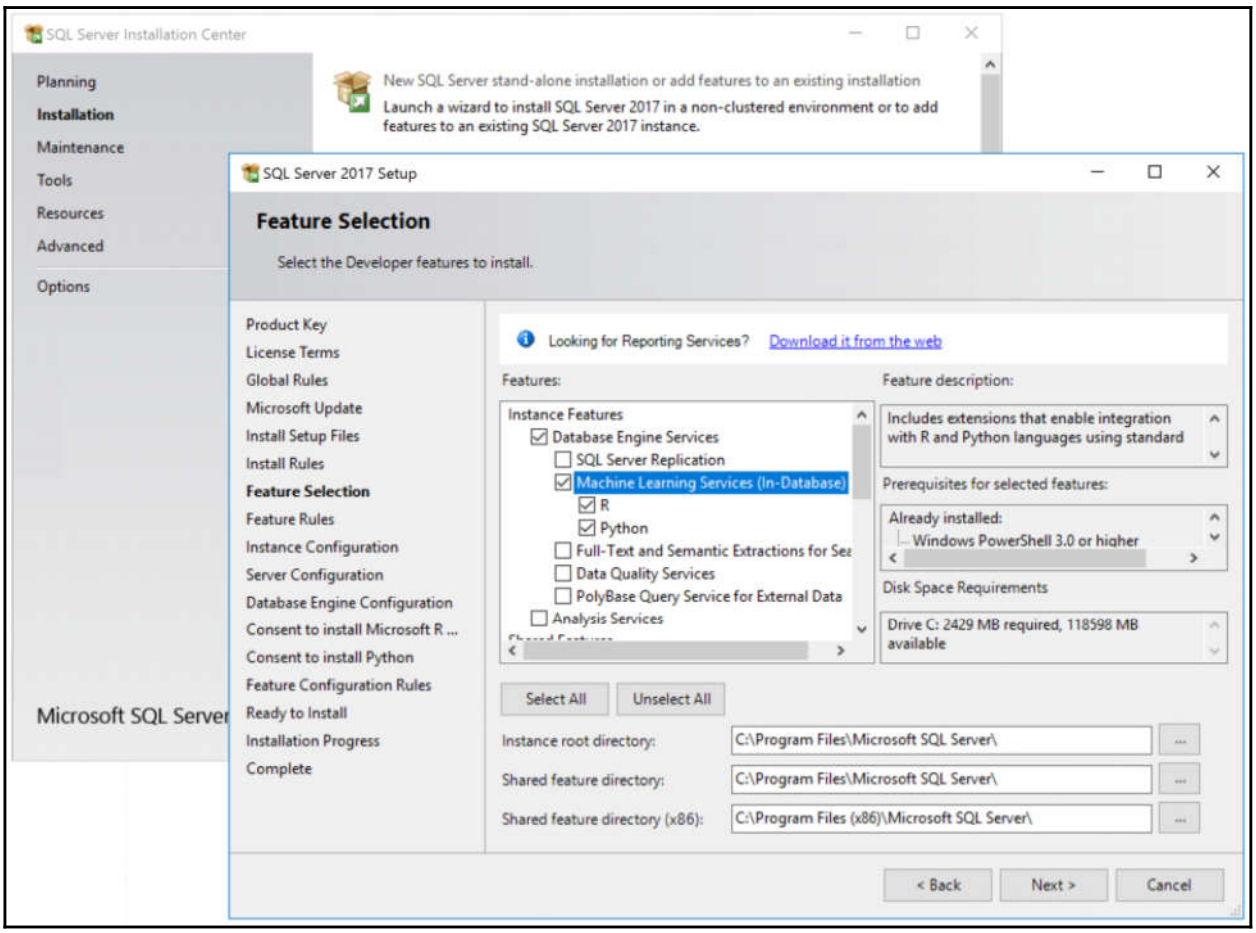
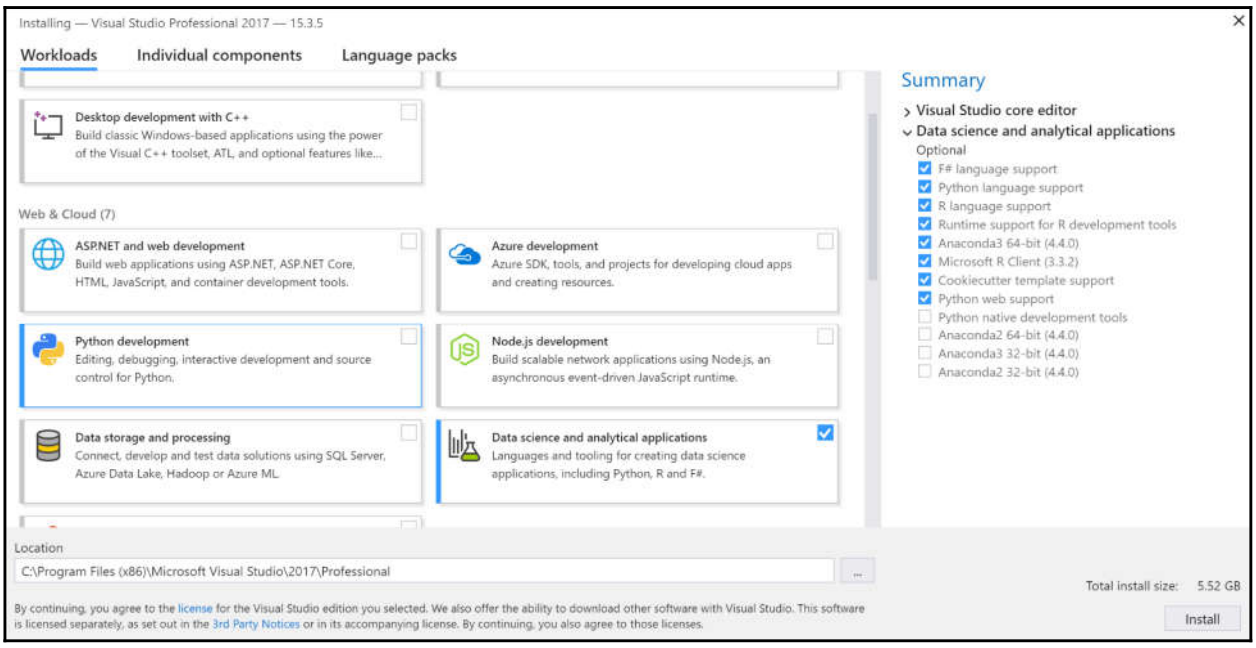
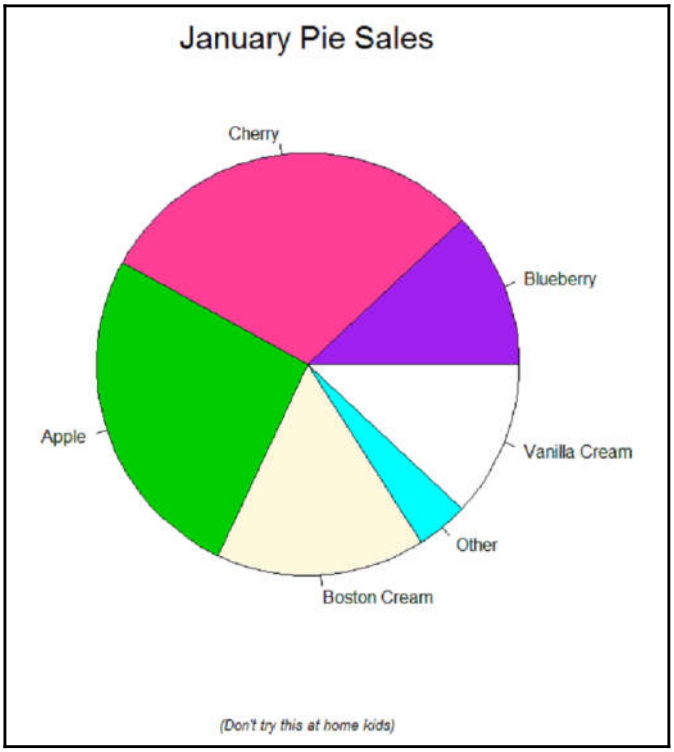
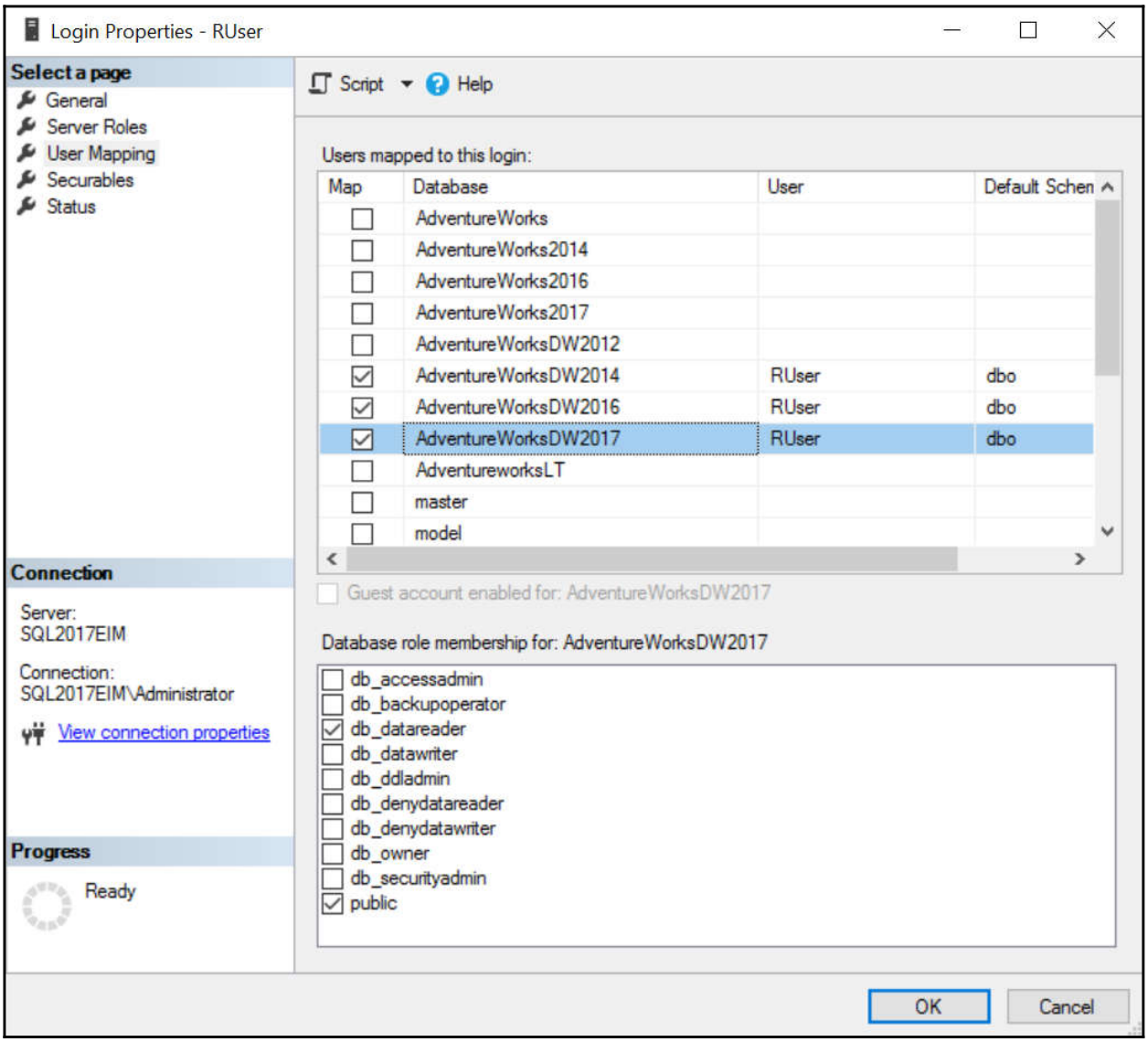
Chapter 1: Writing Queries with T-SQL



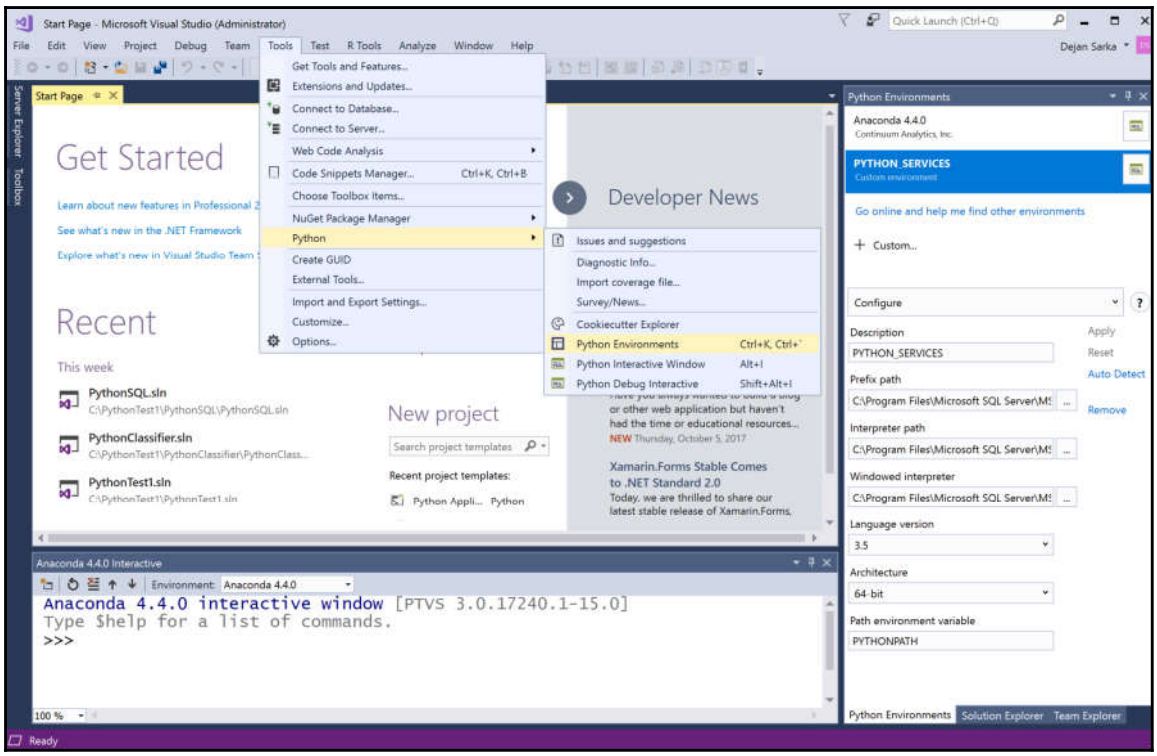
Chapter 2: Introducing R

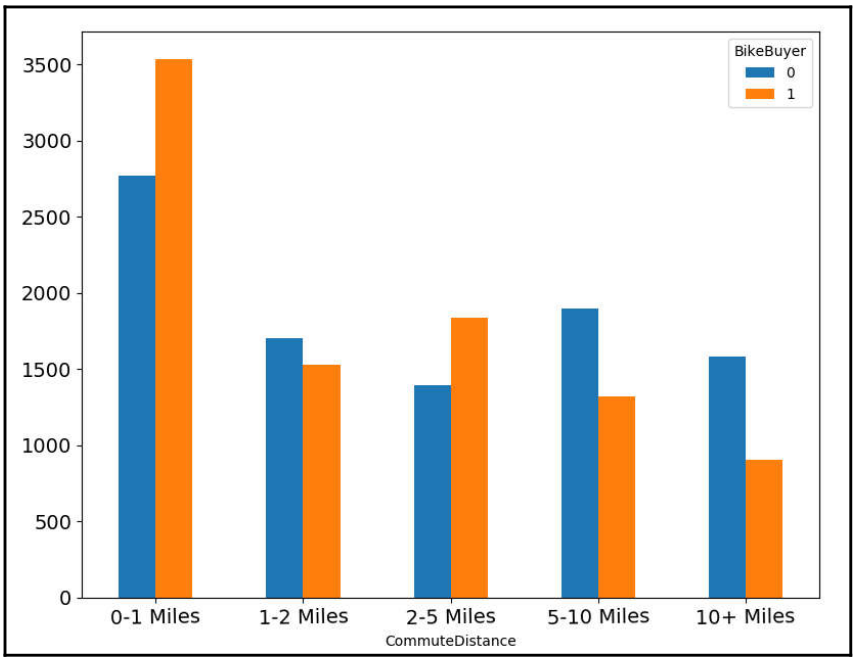


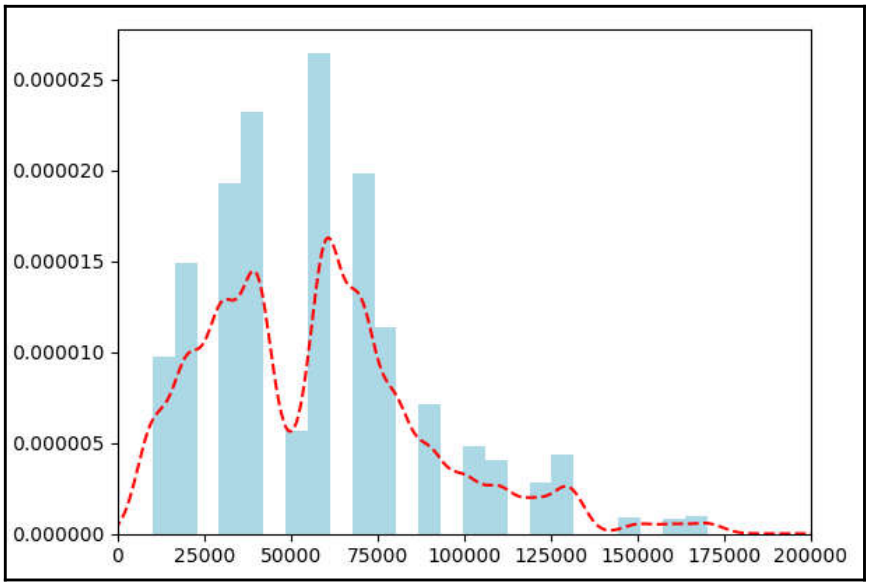


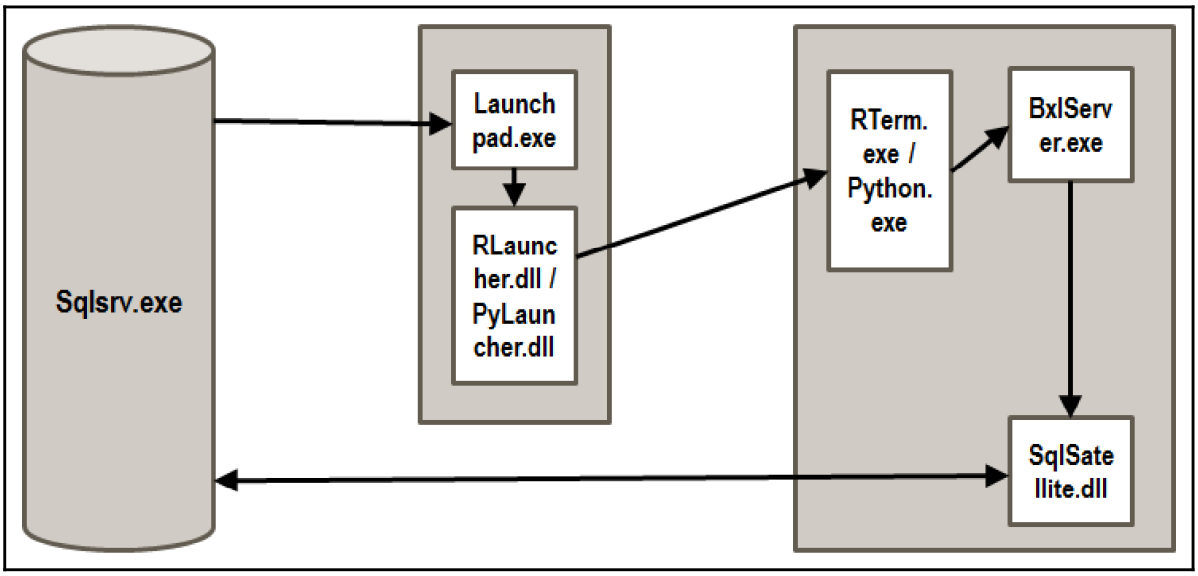

Login Properties - RUser
Select a page
General
Server Roles
User Mapping
Securables
Status
Script
Help
Users mapped to this login:
Map
Database
User
Default Schen
AdventureWorks
AdventureWorks2014
AdventureWorks2016
AdventureWorks2017
AdventureWorksDW2012
AdventureWorksDW2014
RUser
dbo
AdventureWorksDW2016
RUser
dbo
AdventureWorksDW2017
RUser
dbo
AdventureworksLT
master
model
Guest account enabled for: AdventureWorksDW2017
Database role membership for: AdventureWorksDW2017
db_accessadmin
db_backupoperator
db_datareader
db_datawriter
db_ddladmin
db_denydatareader
db_denydatawriter
db_owner
db_securityadmin
public
Connection
Server:
SQL2017EIM
Connection:
SQL2017EIM\Administrator
View connection properties
Progress
Ready
OK
Cancel

Chapter 3: Getting Familiar with Python

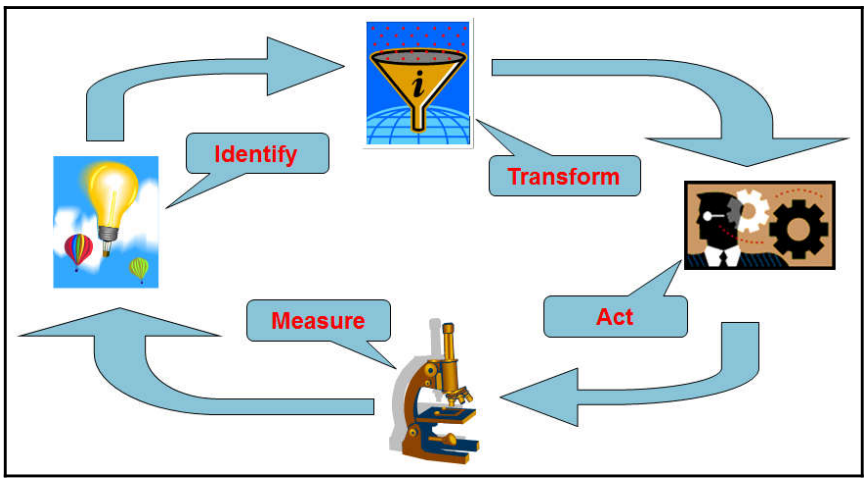


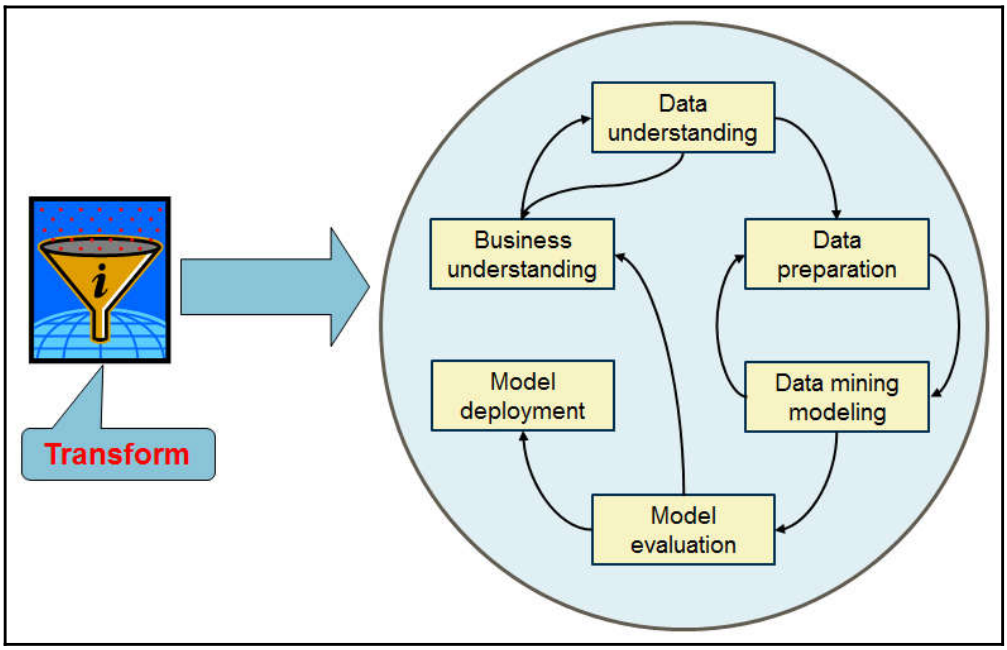

BikeBuyer
0
1
3500
3000
2500
2000
1500
1000
500
0
0-1 Miles
1-2 Miles
2-5 Miles
5-10 Miles
10+ Miles
CommuteDistance


0.000025
0.000020
0.000015
0.000010
0.000005
0.000000
0
25000
50000
75000
100000
125000
150000
175000
200000


Sqlsrv.exe
Launch pad.exe
RLaunc her.dll / PyLaun cher.dll
RTerm. exe / Python. exe
BxlServ er.exe
SqlSate llite.dll

Chapter 4: Data Overview





$$\mu = 1/n * \sum_{i=1}^{n} v_i$$

$$R = v_{max} - v_{min}$$

$$\boxed{IQR = Q_3 - Q_1}$$

$$\boxed{Var = 1/(n-1) * \sum_{i=1}^{n} (v_i - \mu)^2}$$

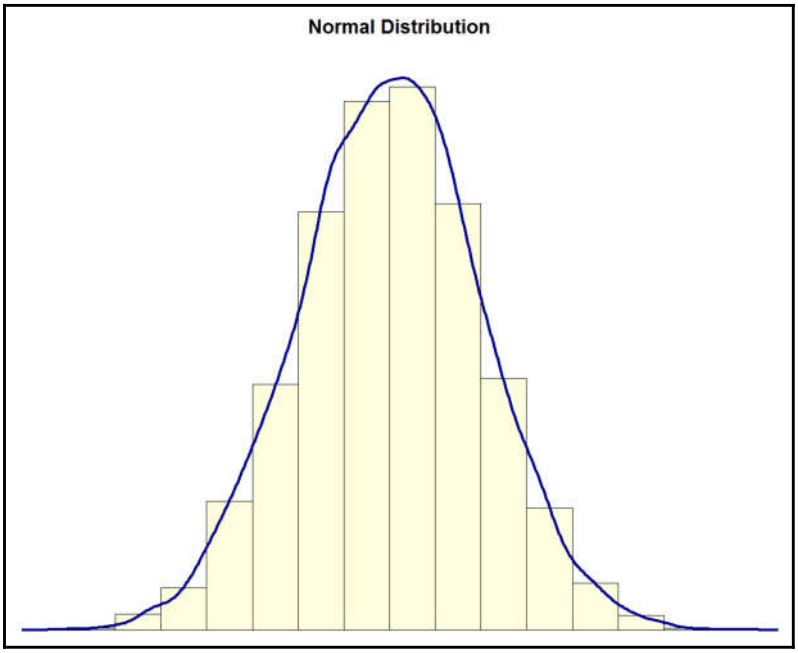
$$\boxed{VarP = 1/n * \sum_{i=1}^{n} (v_i - \mu)^2}$$

$$\boxed{\sigma = \sqrt{Var}}$$

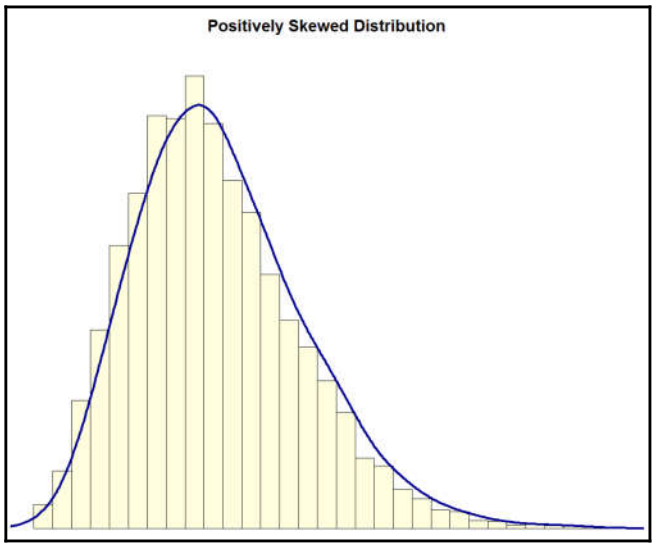
$$\boxed{CV = \sigma/\mu}$$

$$\boxed{f(x, \mu, \sigma) = 1/\sqrt{2\pi\sigma} * e^{-(x-\mu)^2/2\sigma^2}}$$

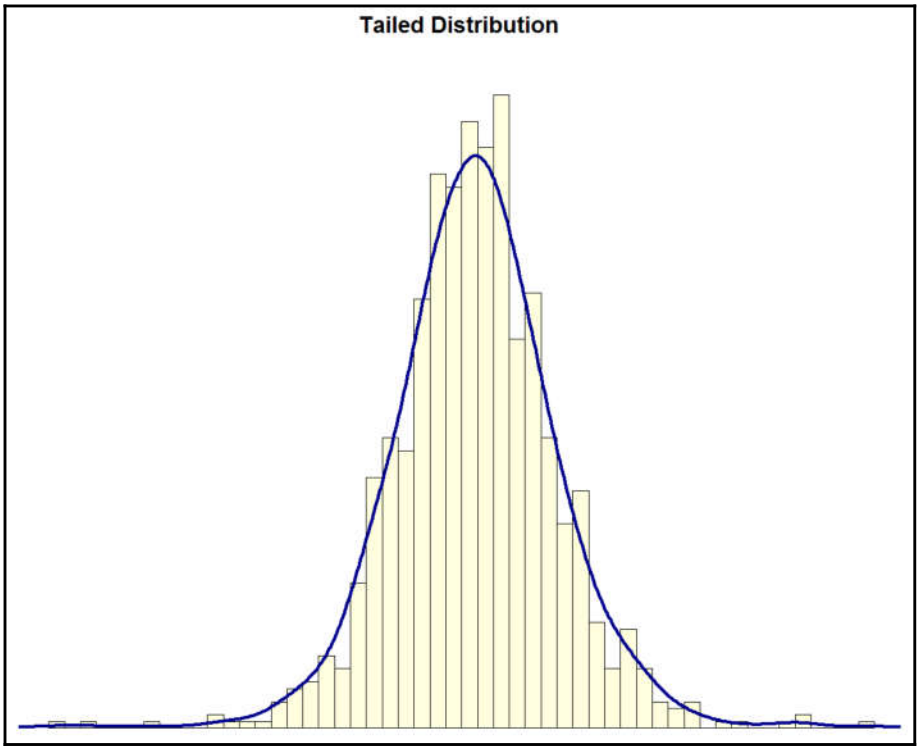
$$\boxed{z = (x - \mu)/\sigma}$$



$$\boxed{Skew = n/((n-1) * (n-2)) * \sum_{i=1}^{n} ((v_i - \mu)/\sigma)^3}$$

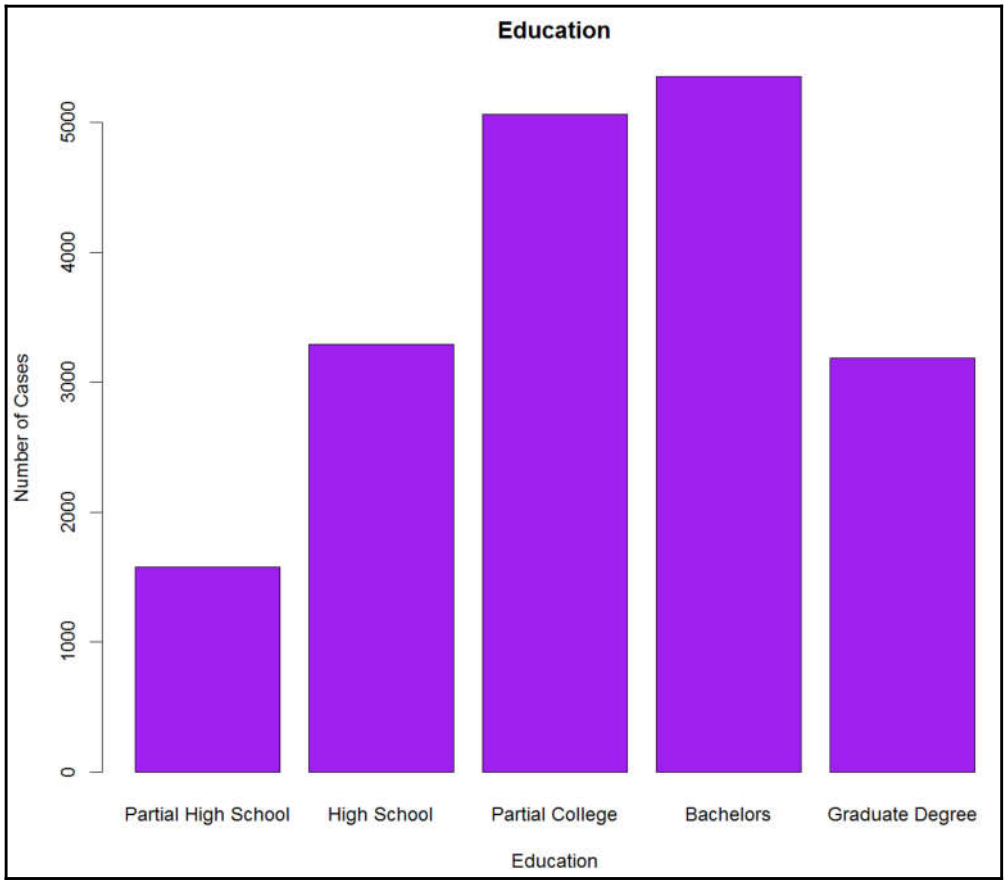


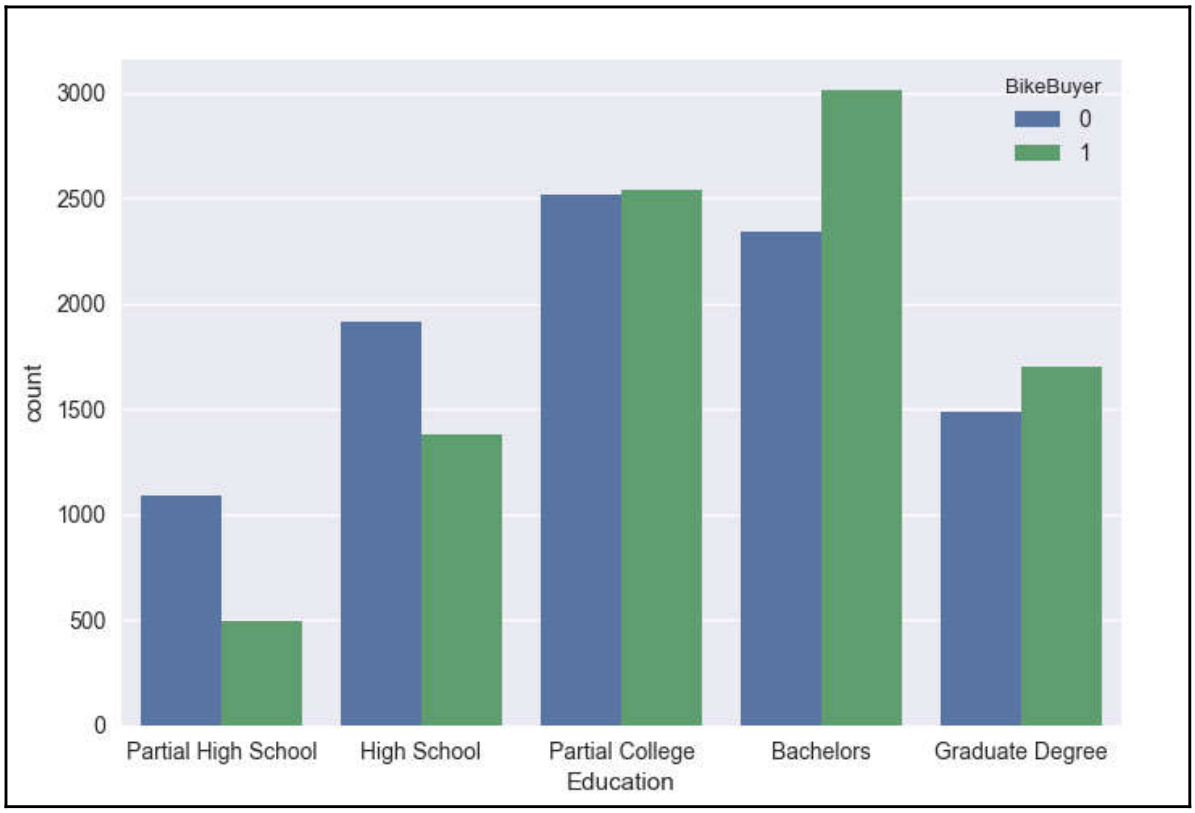
$$Kurt = (n*(n+1))/((n-1)*(n-2)*(n-3))*\sum_{i=1}^{n}((v_i-\mu)/\sigma)^4-(3*(n-1)^2)/((n-2)*(n-3))$$

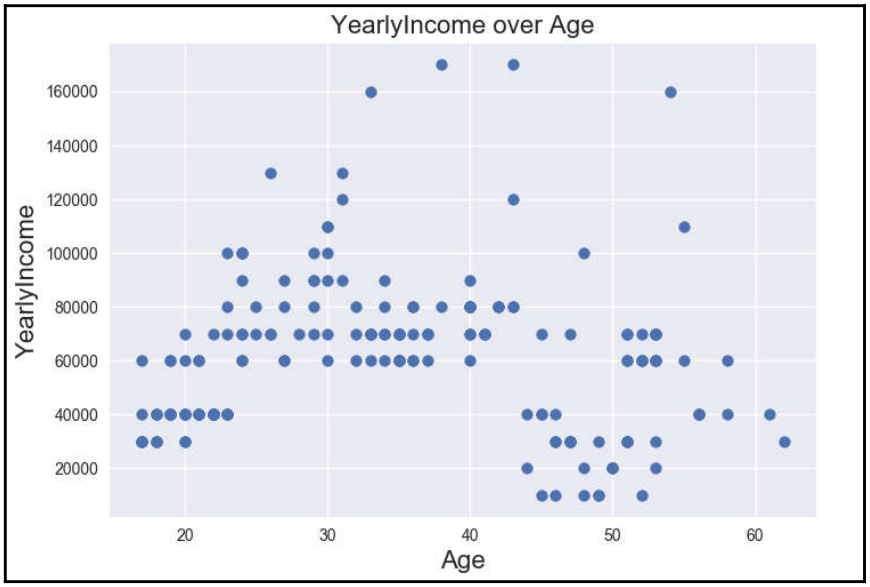


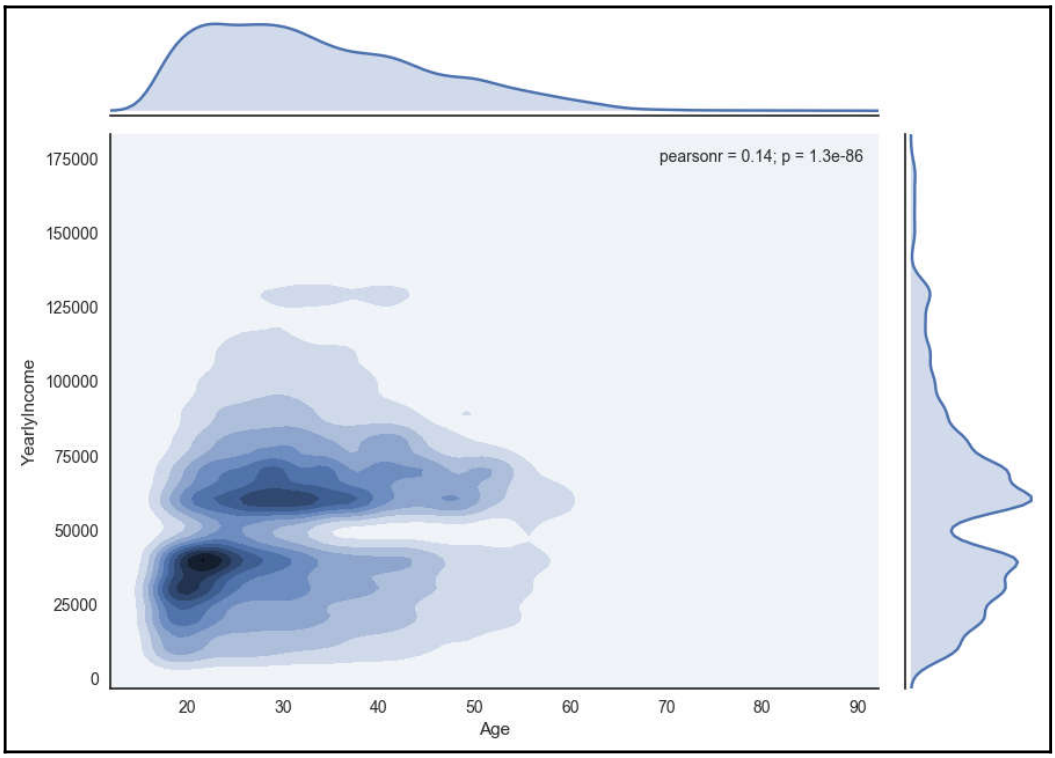
Results | Messages

	CommuteDistance	AbsFreq	CumFreq	AbsPerc	CumPerc	Histogram
1	1 - 0-1 Miles	6310	6310	34	34	**********
2	2 - 1-2 Miles	3232	9542	17	51	**********
3	3 - 2-5 Miles	3234	12776	17	68	**********
4	4 - 5-10 Miles	3214	15990	17	85	**********
5	5 - 10+ Miles	2494	18484	13	98	**********

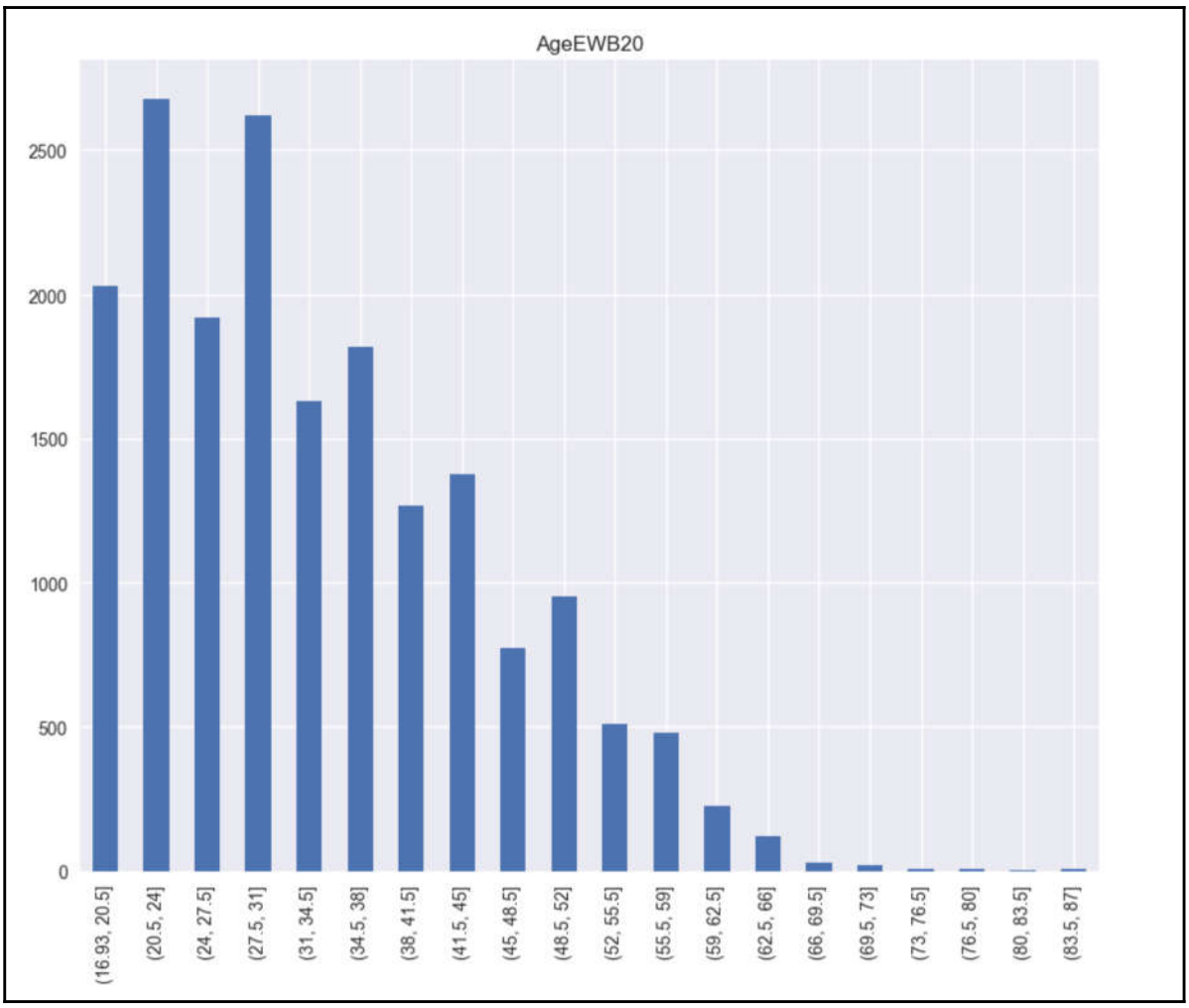


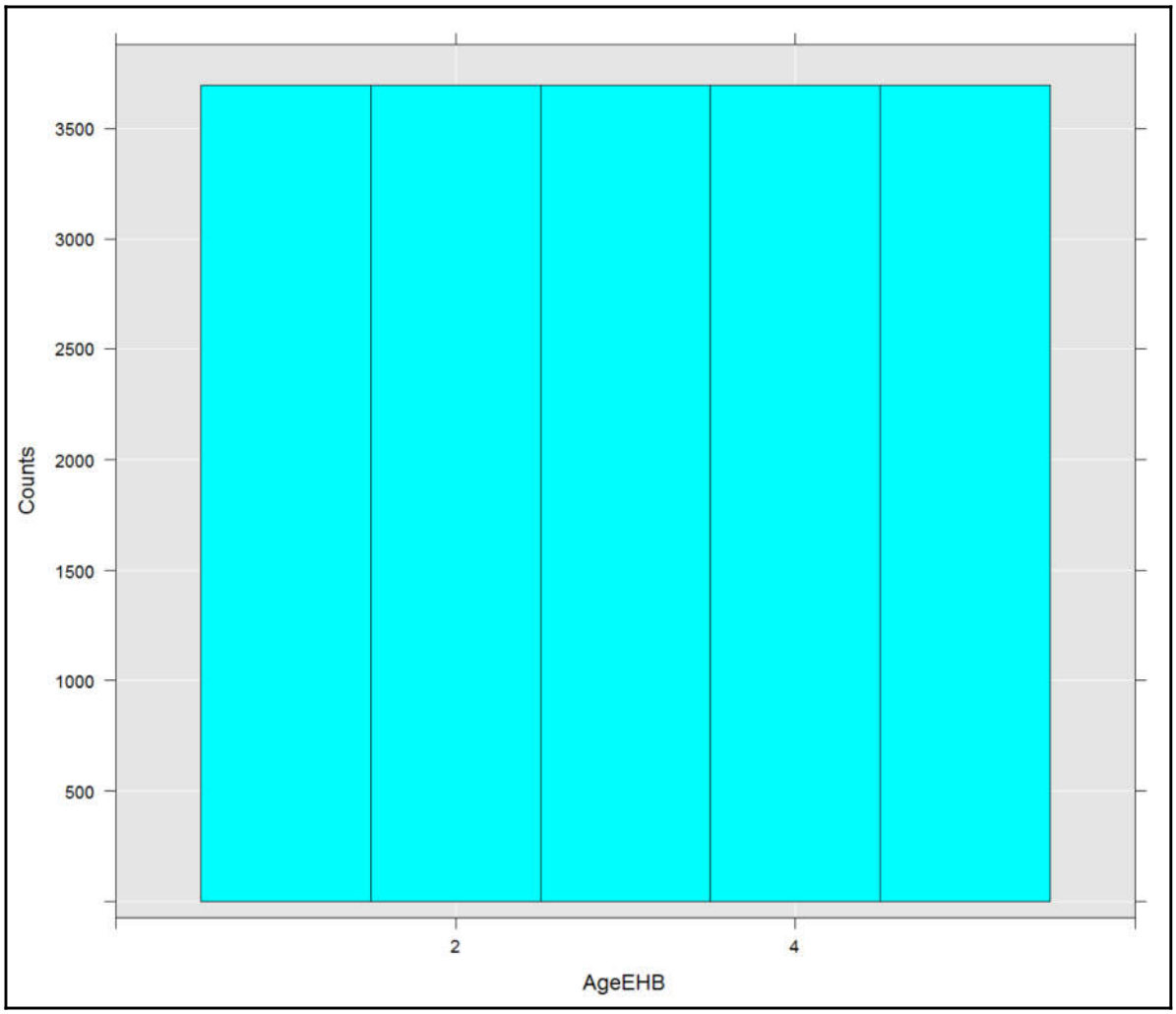

BikeBuyer
0
1
count
3000
2500
2000
1500
1000
500
0
Partial High School
High School
Partial College
Bachelors
Graduate Degree
Education

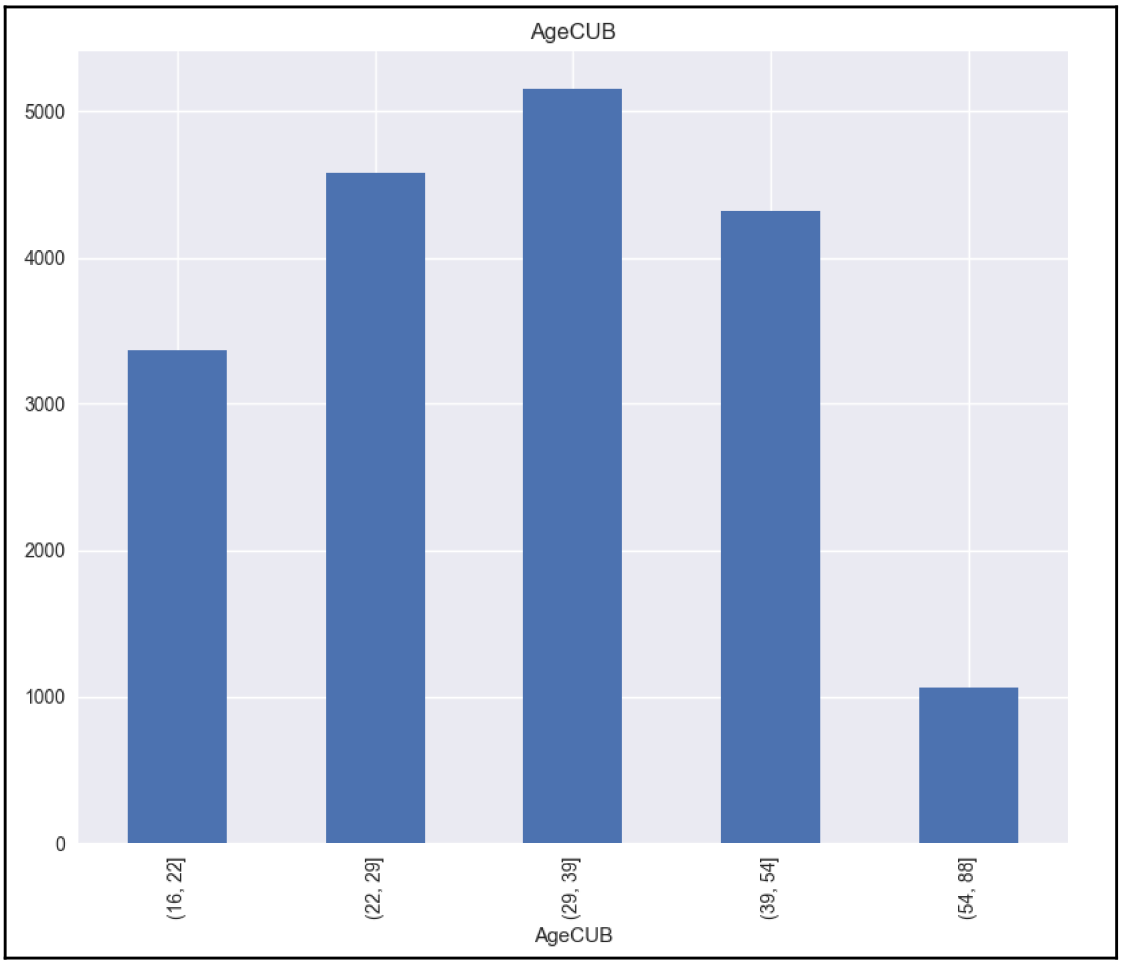

YearlyIncome over Age
YearlyIncome
160000
140000
120000
100000
80000
60000
40000
20000
20
30
40
50
60
Age


pearsonr = 0.14; p = 1.3e-86
YearlyIncome
175000
150000
125000
100000
75000
50000
25000
0
20
30
40
50
60
70
80
90
Age

Chapter 5: Data Preparation



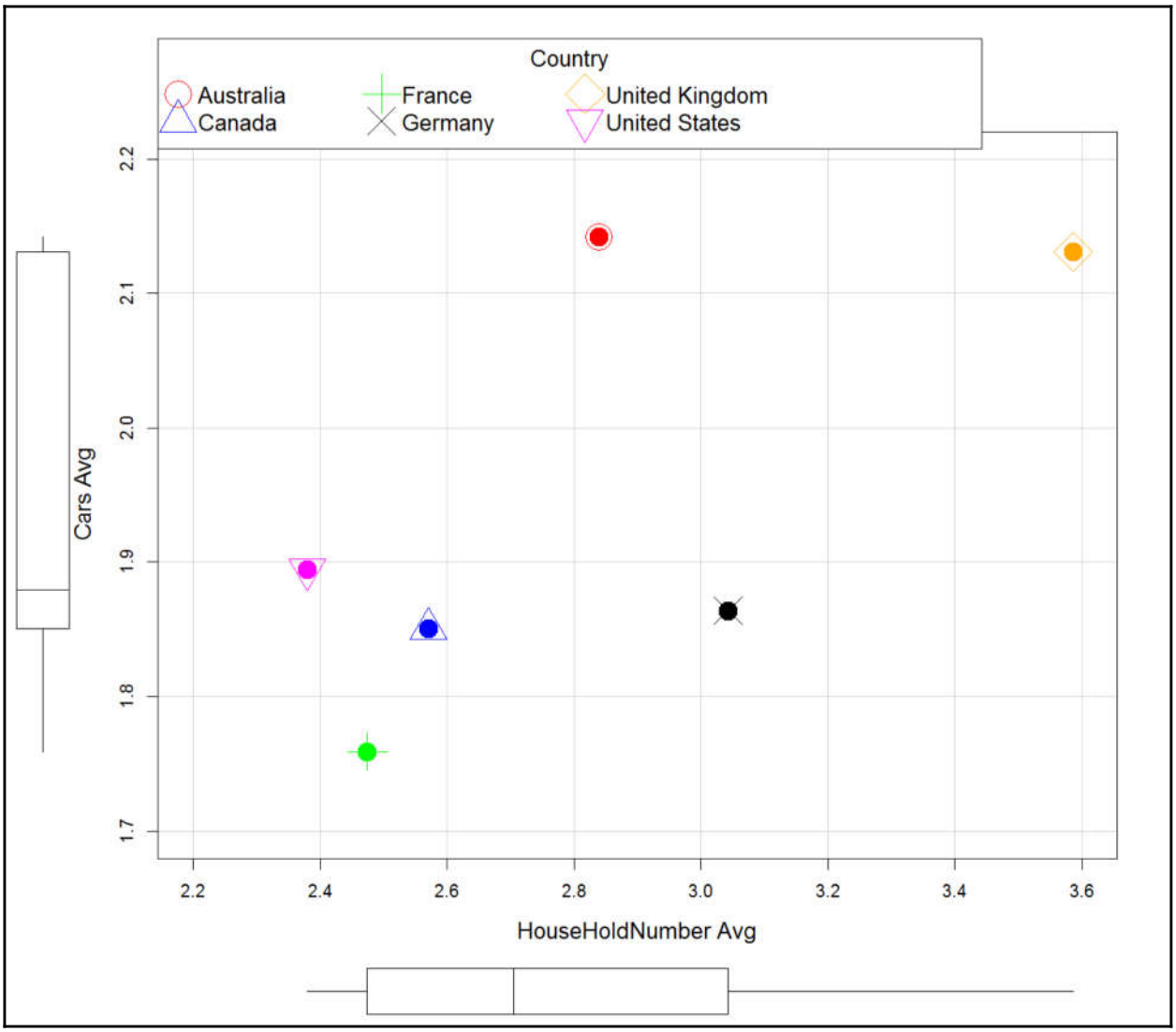

Counts
3500
3000
2500
2000
1500
1000
500
2
4
AgeEHB



$$\boxed{I(x) = -P(x_i) * LOG_2(P(x_i))}$$

$$\boxed{H(x) = -\sum_{i=1}^{n} P(x_i) * LOG_2(P(x_i))}$$

$$\boxed{H_{max,n} = LOG_2(n)}$$

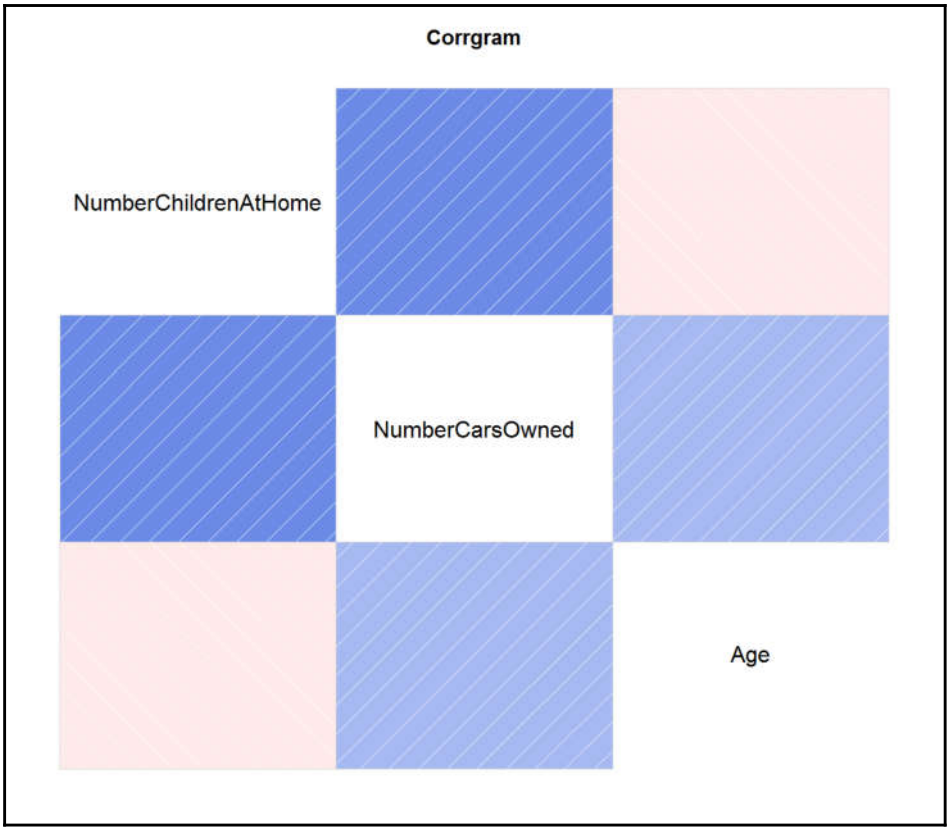

Country
Australia
Canada
France
Germany
United Kingdom
United States
Cars Avg
1.7
1.8
1.9
2.0
2.1
2.2
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6
HouseHoldNumber Avg

Chapter 6:
Intermediate Statistics and Graphs

$$\boxed{CoVar(X,Y) = 1/n * \sum_{i=1}^{n}(X_i - \mu(X)) * (Y_i - \mu(Y))}$$

$$\boxed{Correl(X,Y) = Covar(X,Y))/((\sigma(X) * \sigma(Y))}$$

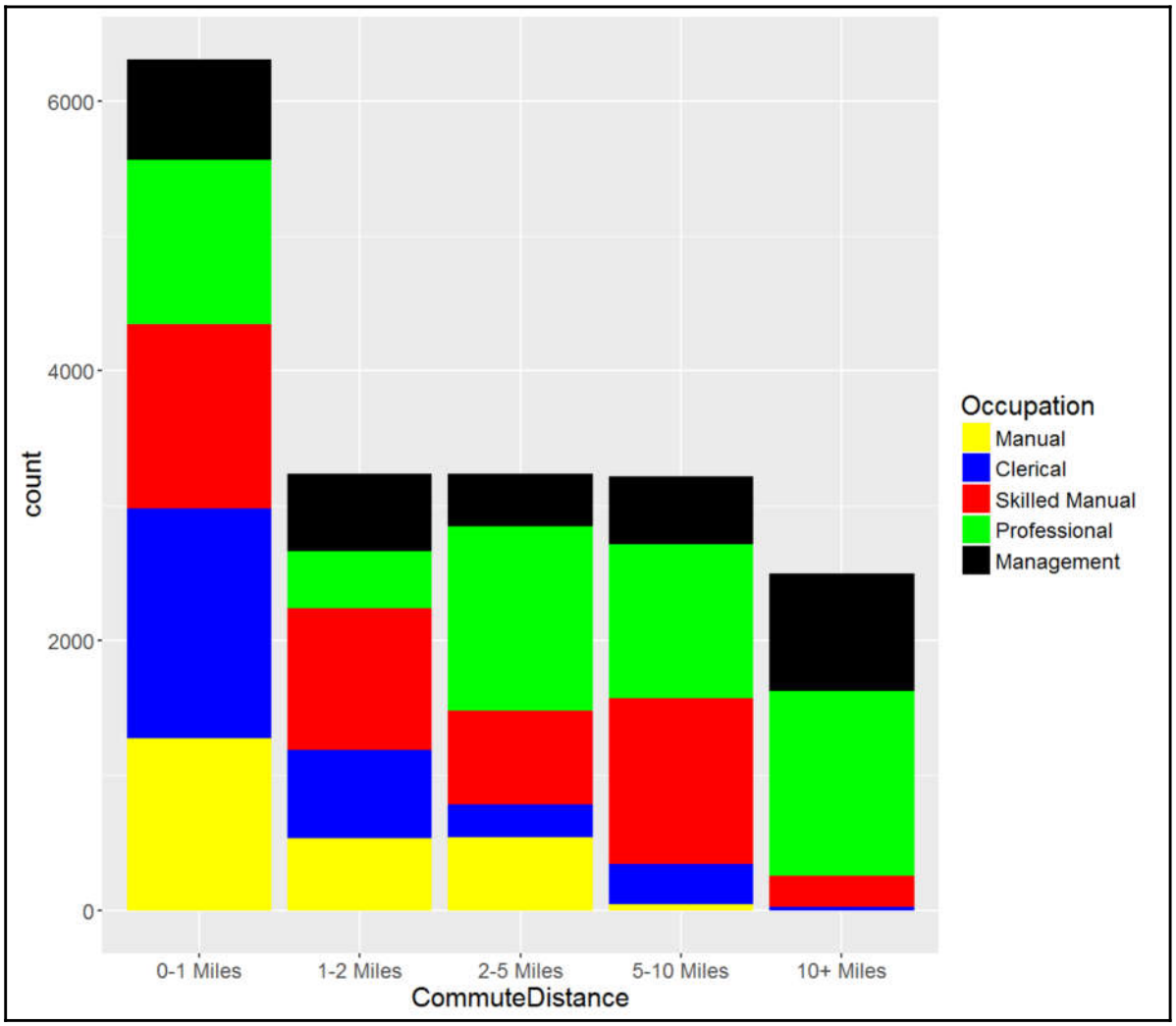
$$\boxed{CD(X,Y) = Correl(X,Y)^2}$$

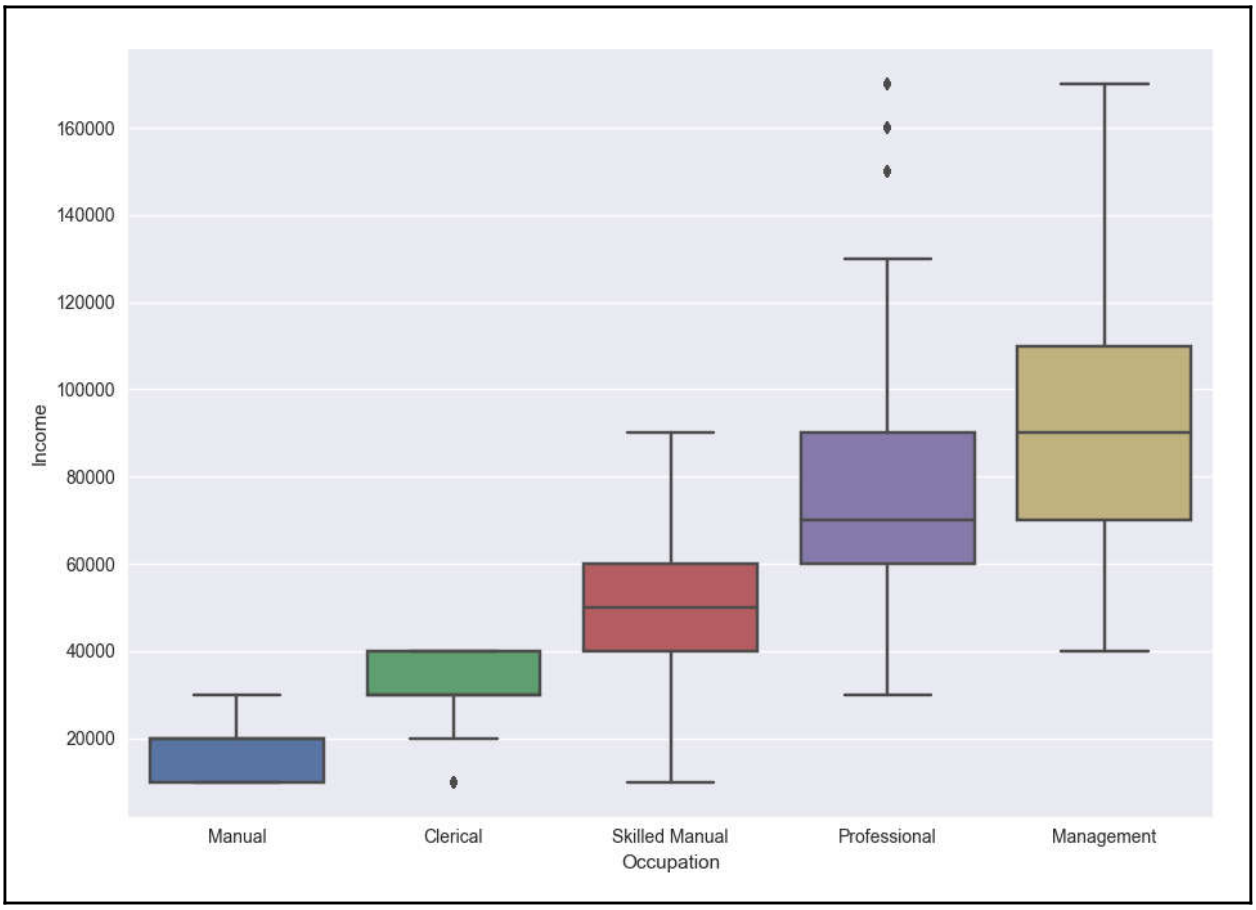


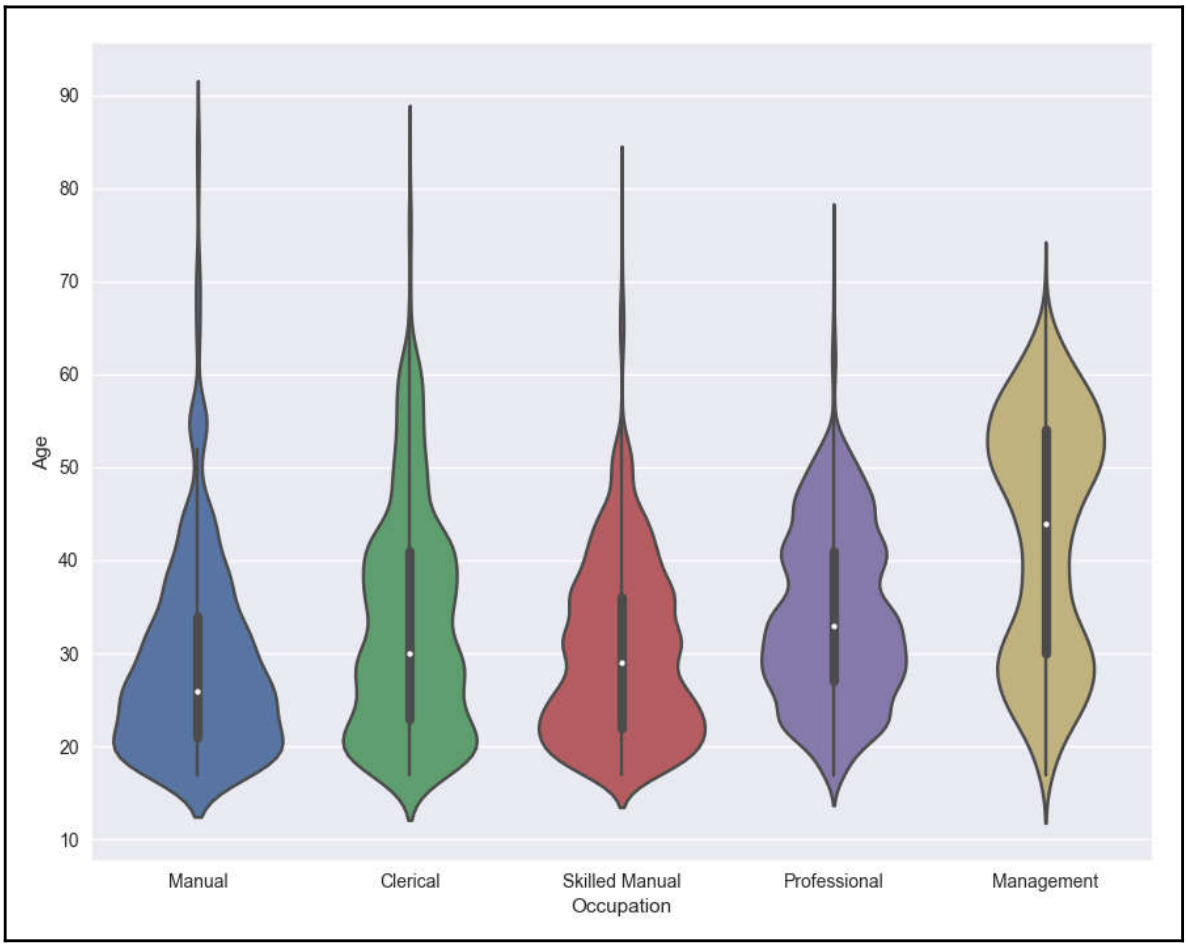
$$\boxed{\chi^2 = 1/n * \sum_{i=1}^{n}(O - E)^2/E}$$

$$\boxed{DF = (C-1) * (R-1)}$$

DF	***Chi-squared Value***										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.56	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
Probability	***0.95***	***0.90***	***0.80***	***0.70***	***0.50***	***0.30***	***0.20***	***0.10***	***0.05***	***0.01***	***0.001***
	Not significant								***Significant***		


6000
4000
2000
0
count
0-1 Miles
1-2 Miles
2-5 Miles
5-10 Miles
10+ Miles
CommuteDistance
Occupation
Manual
Clerical
Skilled Manual
Professional
Management

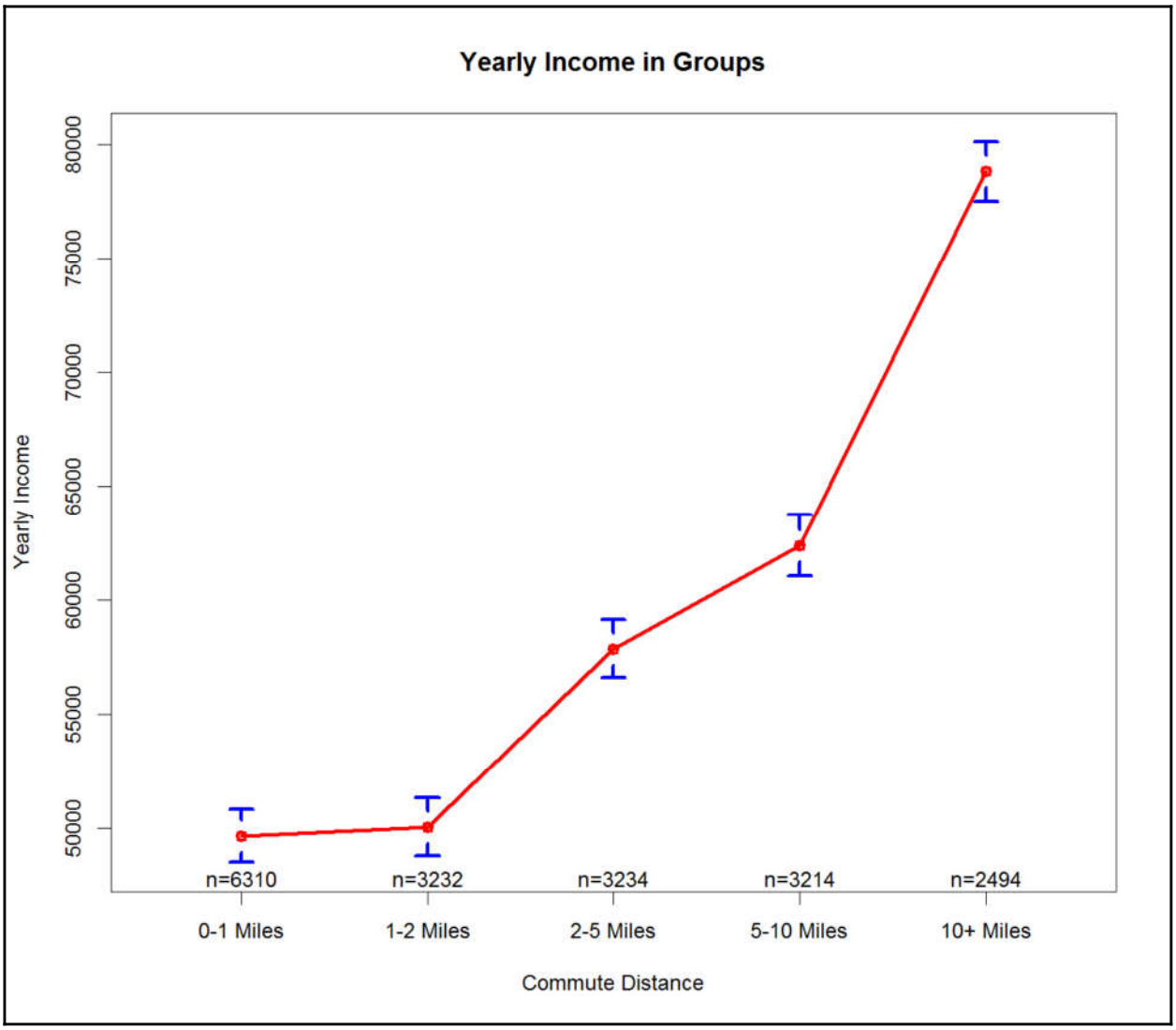

160000
140000
120000
100000
80000
60000
40000
20000
Income
Manual
Clerical
Skilled Manual
Professional
Management
Occupation



$$\boxed{MS_A = SS_A/DF_A; SS_A = \sum_{i=1}^{a} n_i * (\mu_i - \mu)^2; DF_A = (a-1)}$$

$$\boxed{MS_E = SS_E/DF_E; SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (v_{ij} - \mu_i)^2; DF_E = \sum_{i=1}^{a} (n_i - 1)}$$

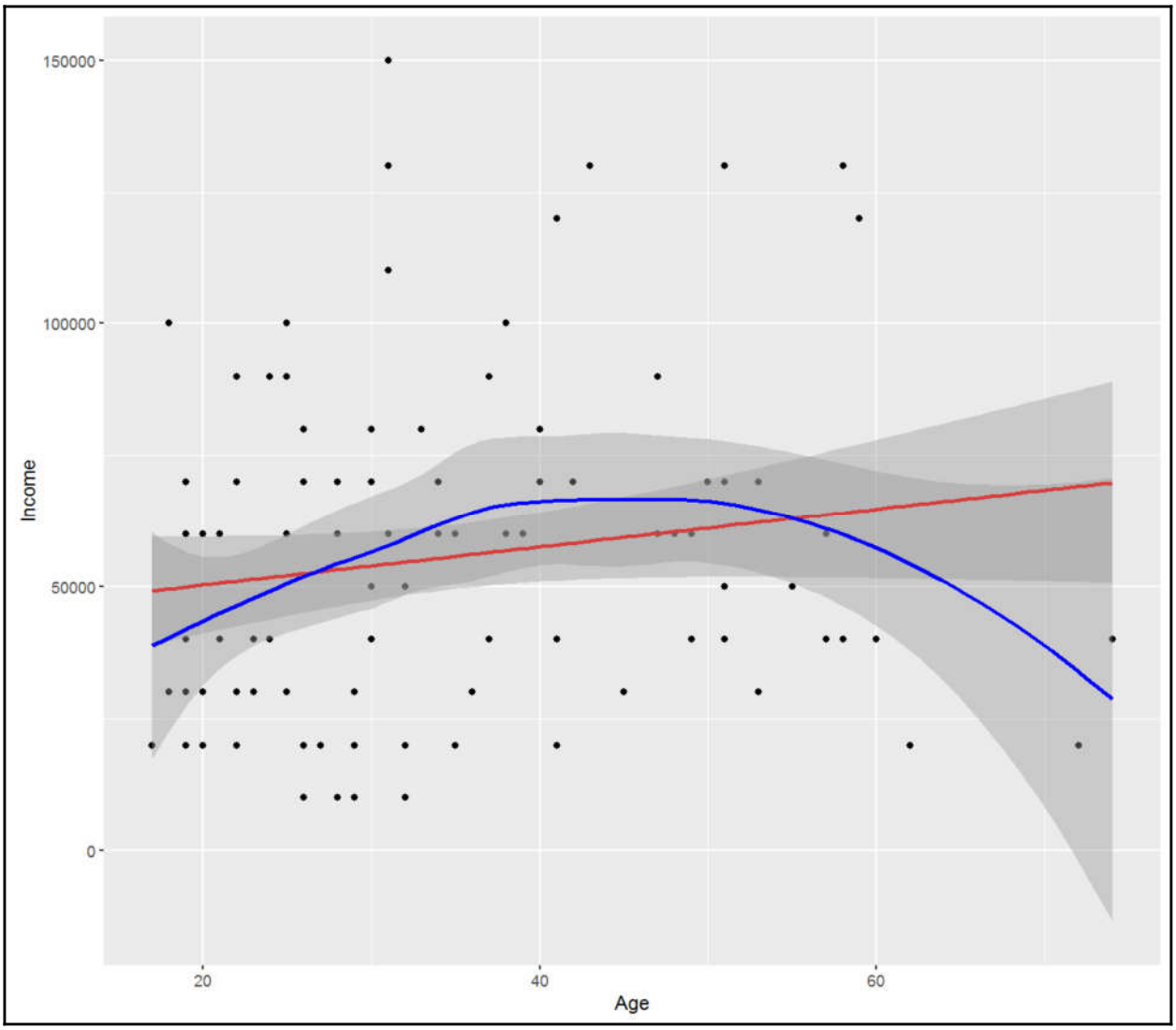
$$\boxed{F = MS_A/MS_E}$$



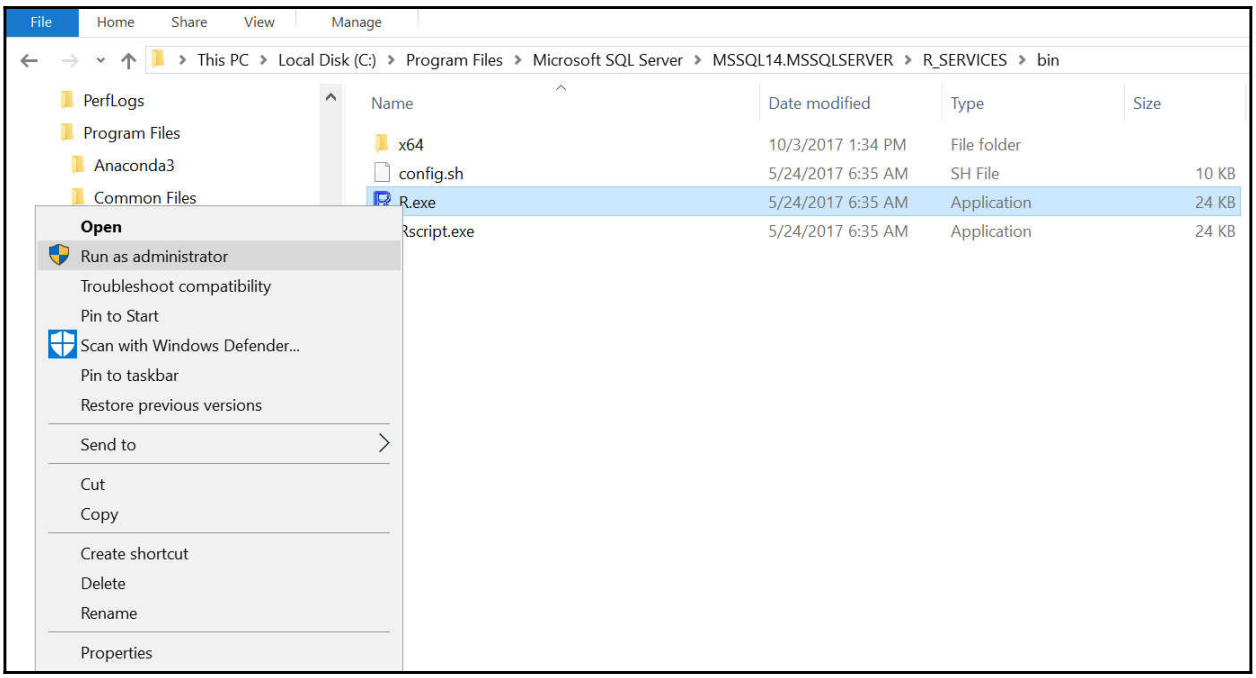
$$\boxed{Y = a + bX}$$

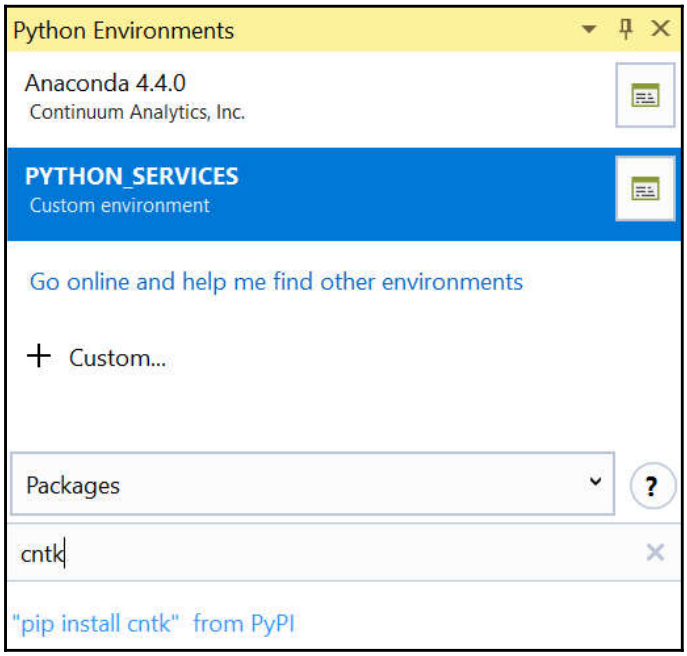
$$\boxed{Slope(Y) = (\sum_{i=1}^{n}(X_i - \mu(X)) * (Y_i - \mu(Y)))/(\sum_{i=1}^{n}(X_i - \mu(X))^2)}$$

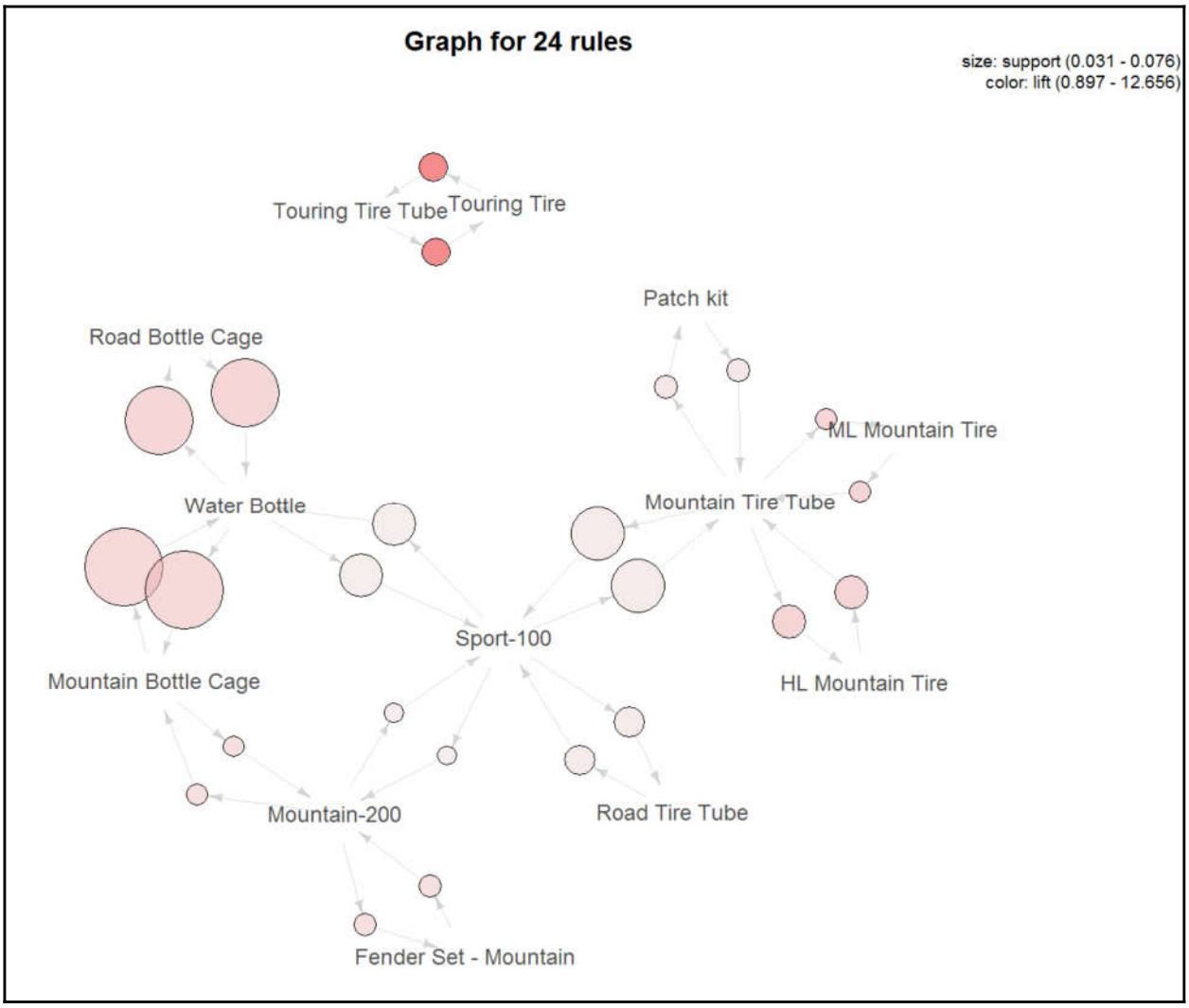
$$\boxed{Intercept(Y) = \mu(Y) - Slope(Y) * \mu(X)}$$

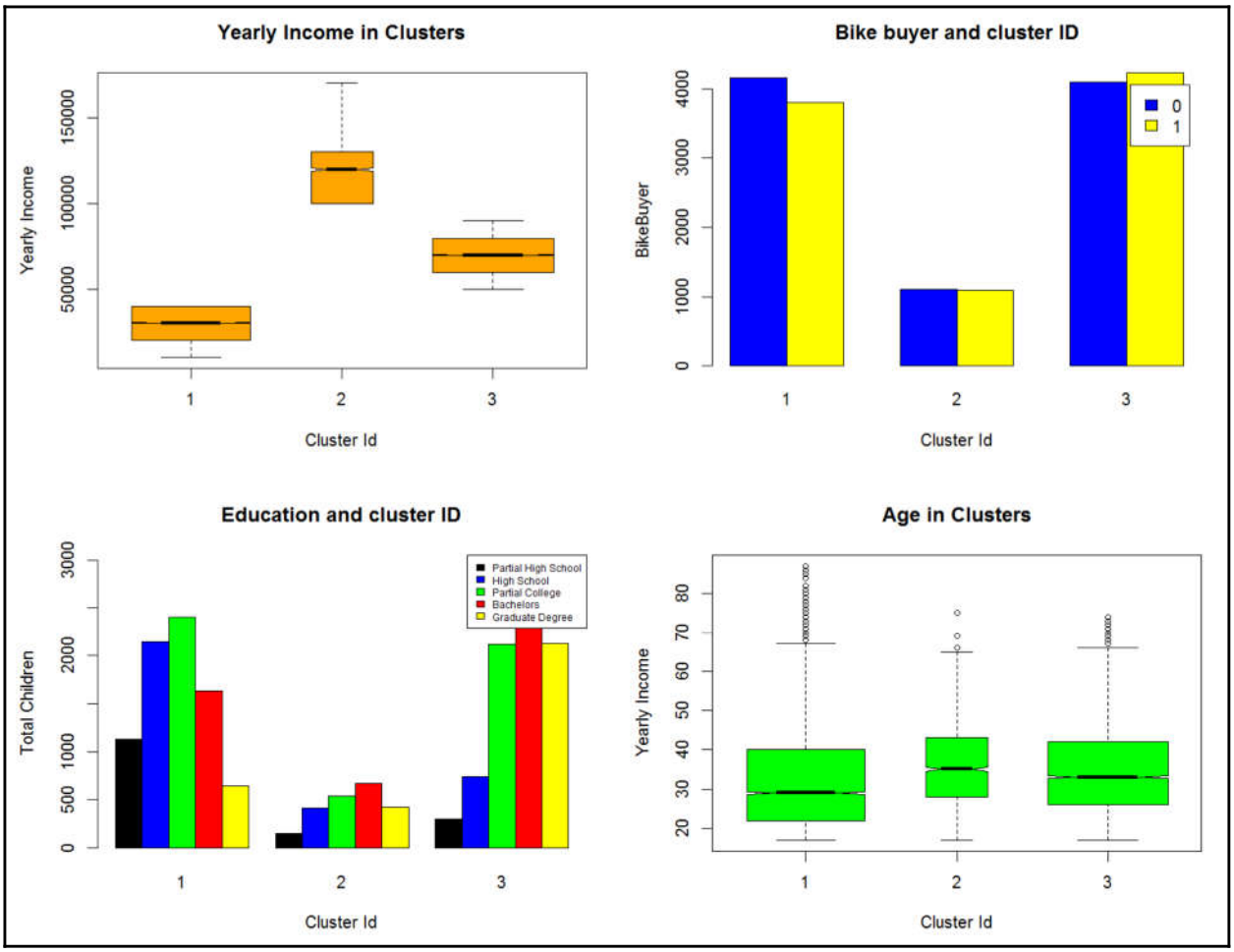

150000
100000
50000
0
Income
20
40
60
Age

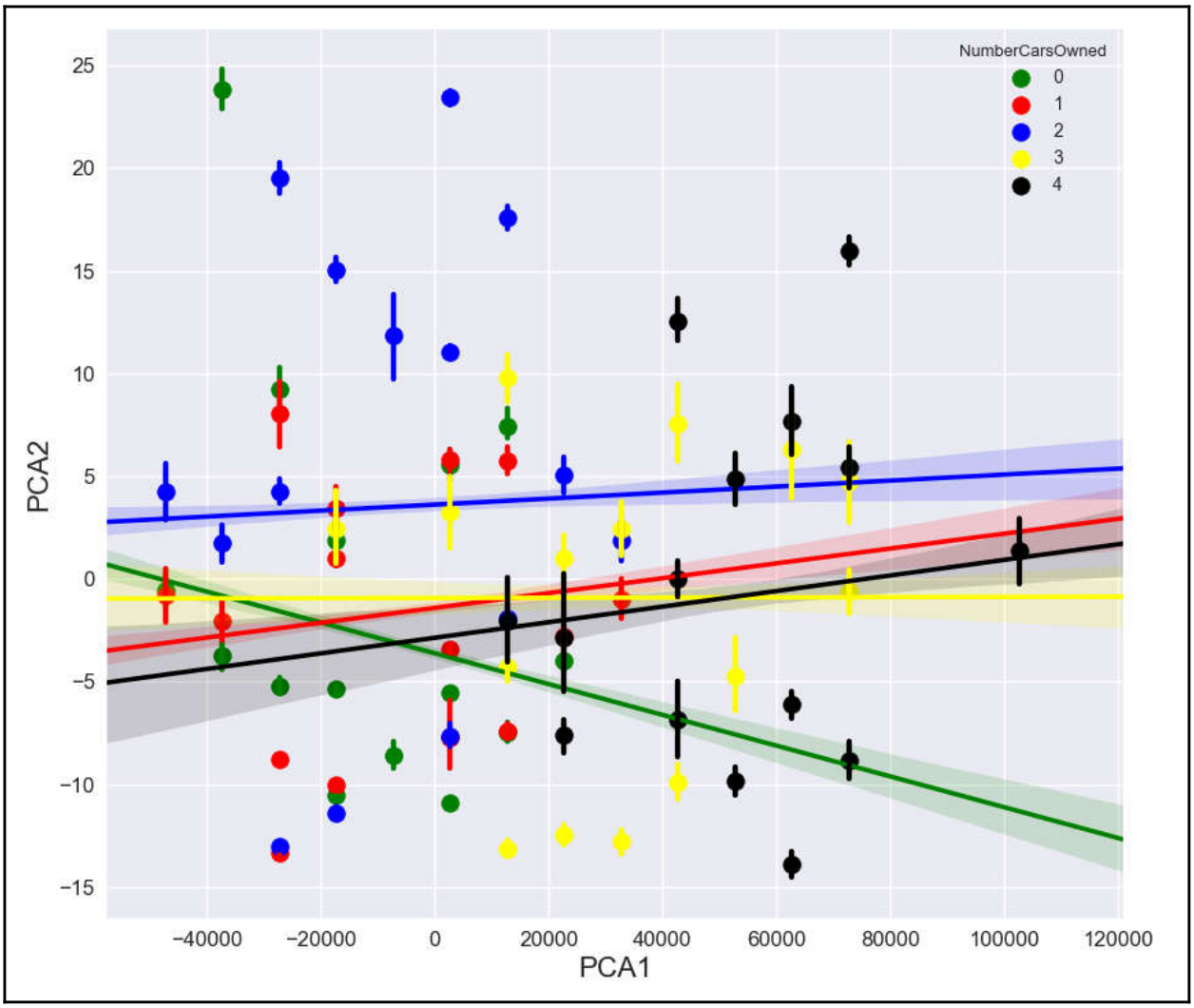
Chapter 7: Unsupervised Machine Learning

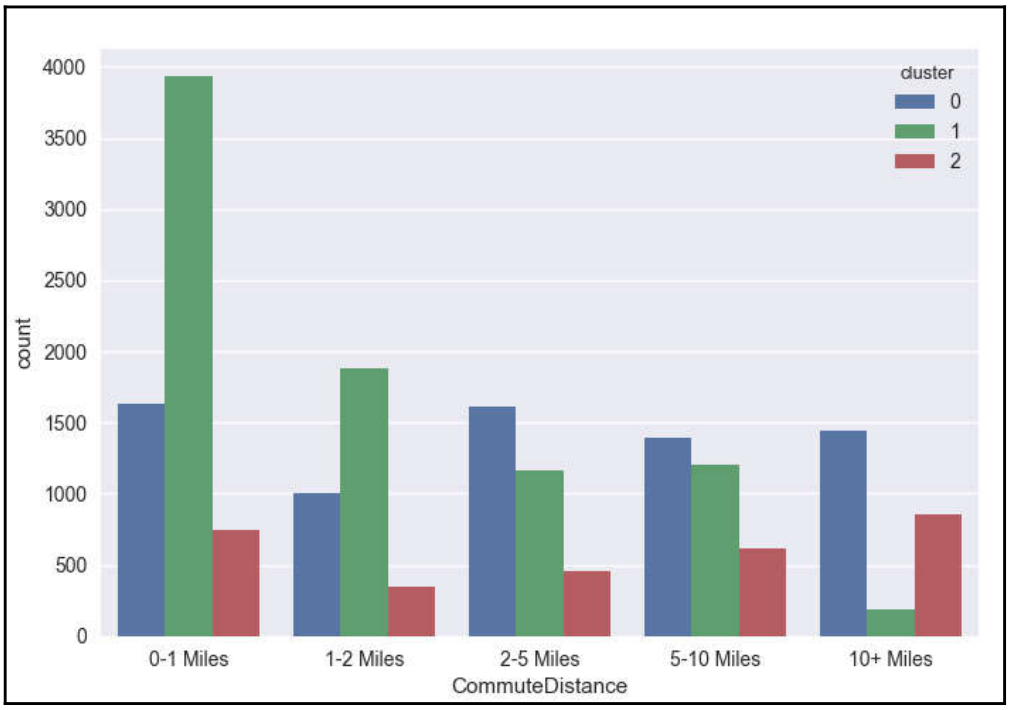


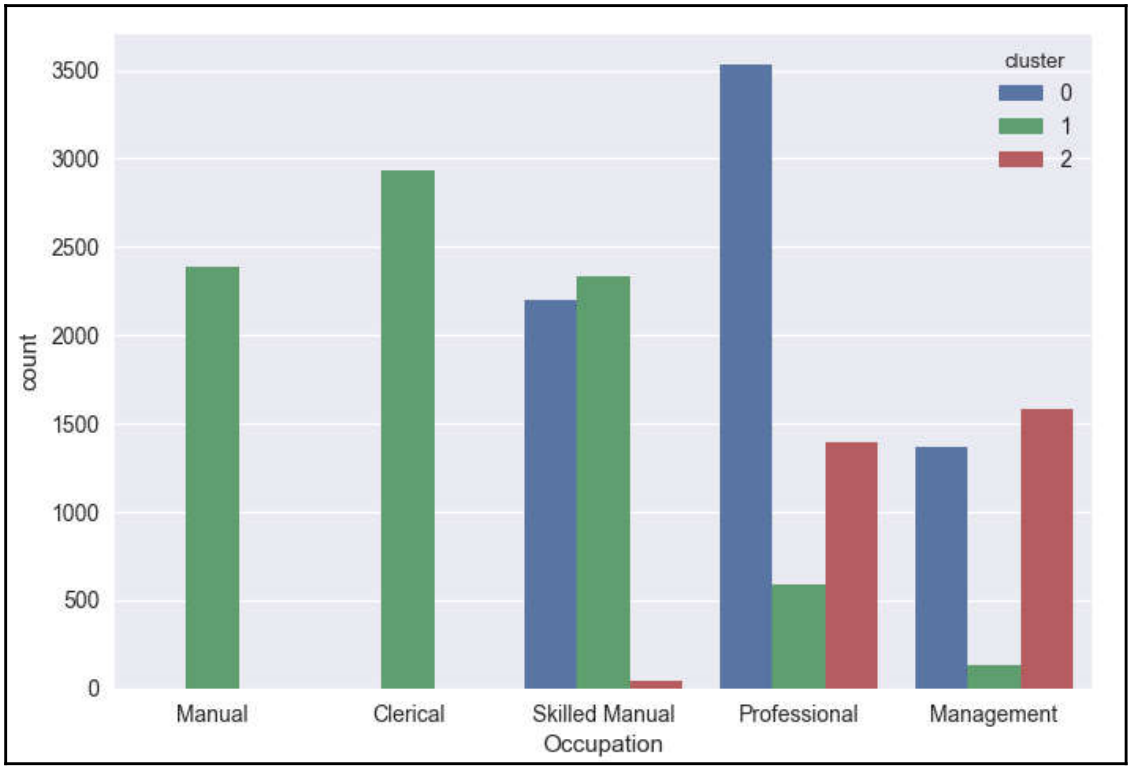


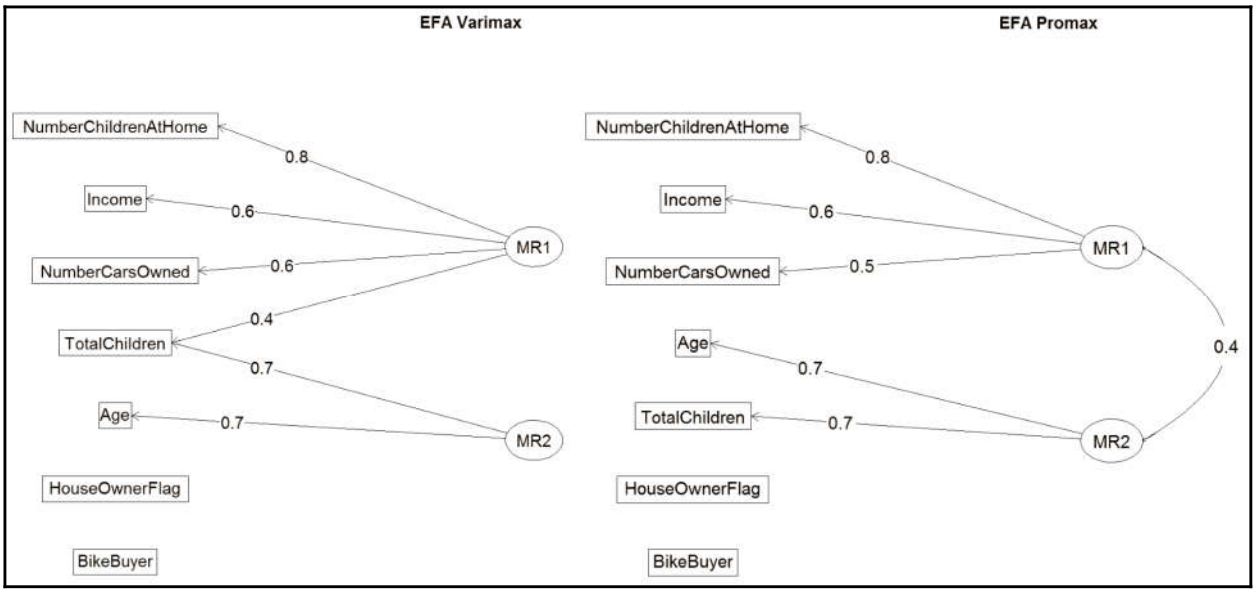

Graph for 24 rules
size: support (0.031 - 0.076)
color: lift (0.897 - 12.656)
Touring Tire Tube
Touring Tire
Patch kit
Road Bottle Cage
ML Mountain Tire
Water Bottle
Mountain Tire Tube
Sport-100
Mountain Bottle Cage
HL Mountain Tire
Mountain-200
Road Tire Tube
Fender Set - Mountain


Yearly Income in Clusters
Yearly Income
50000
100000
150000
1
2
3
Cluster Id
Bike buyer and cluster ID
BikeBuyer
0
1000
2000
3000
4000
0
1
1
2
3
Cluster Id
Education and cluster ID
Total Children
0
500
1000
2000
3000
Partial High School
High School
Partial College
Bachelors
Graduate Degree
1
2
3
Cluster Id
Age in Clusters
Yearly Income
20
30
40
50
60
70
80
1
2
3
Cluster Id

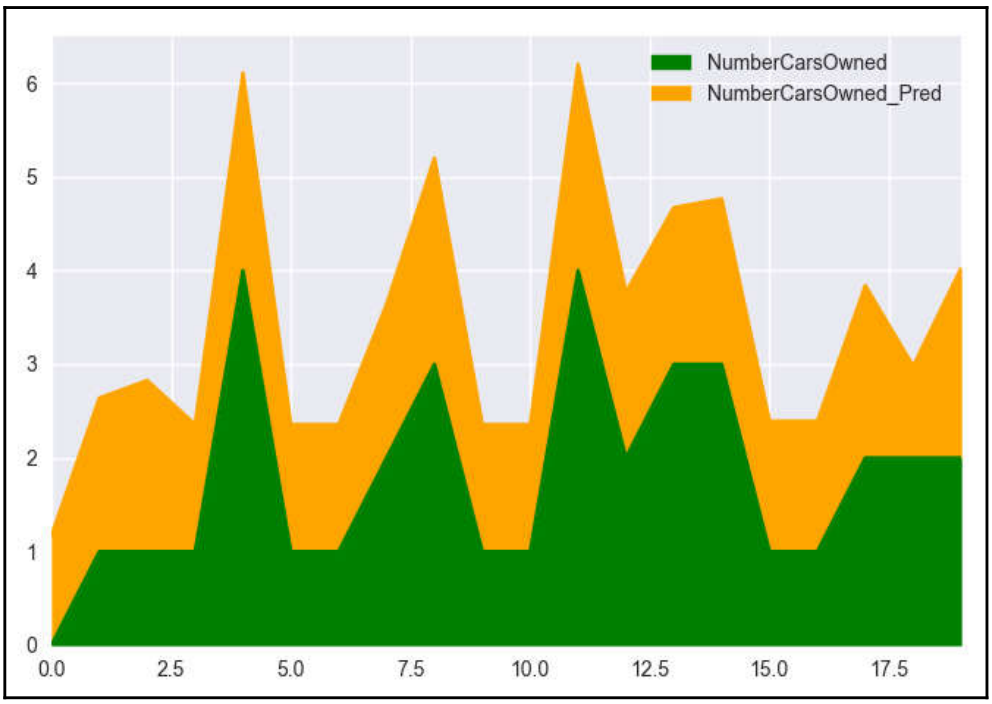

NumberCarsOwned
0
1
2
3
4
PCA2
PCA1
25
20
15
10
5
0
−5
−10
−15
−40000
−20000
0
20000
40000
60000
80000
100000
120000

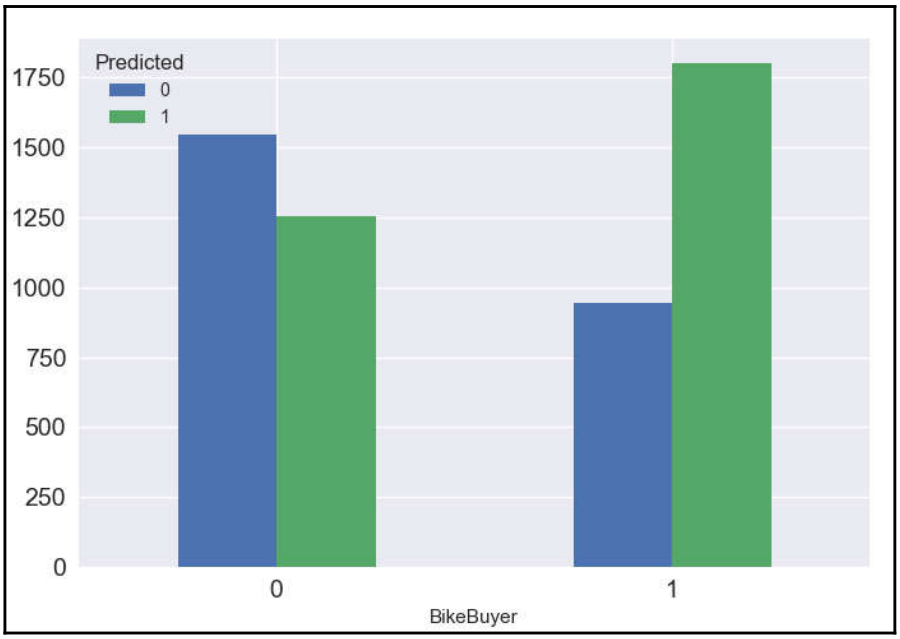

cluster
0
1
2
count
4000
3500
3000
2500
2000
1500
1000
500
0
0-1 Miles
1-2 Miles
2-5 Miles
5-10 Miles
10+ Miles
CommuteDistance

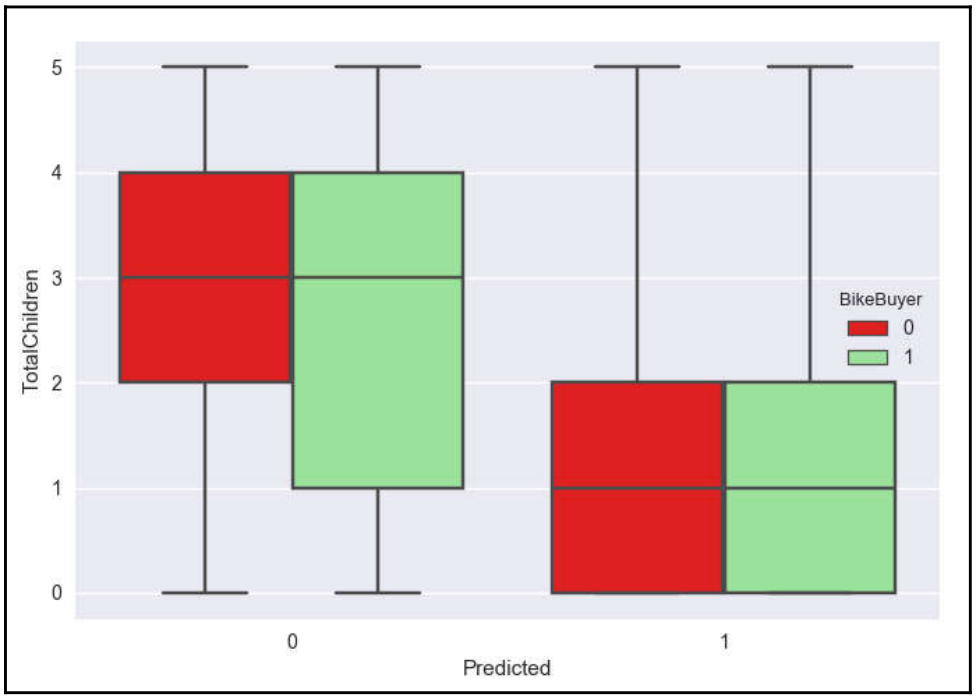

cluster
0
1
2
count
3500
3000
2500
2000
1500
1000
500
0
Manual
Clerical
Skilled Manual
Professional
Management
Occupation

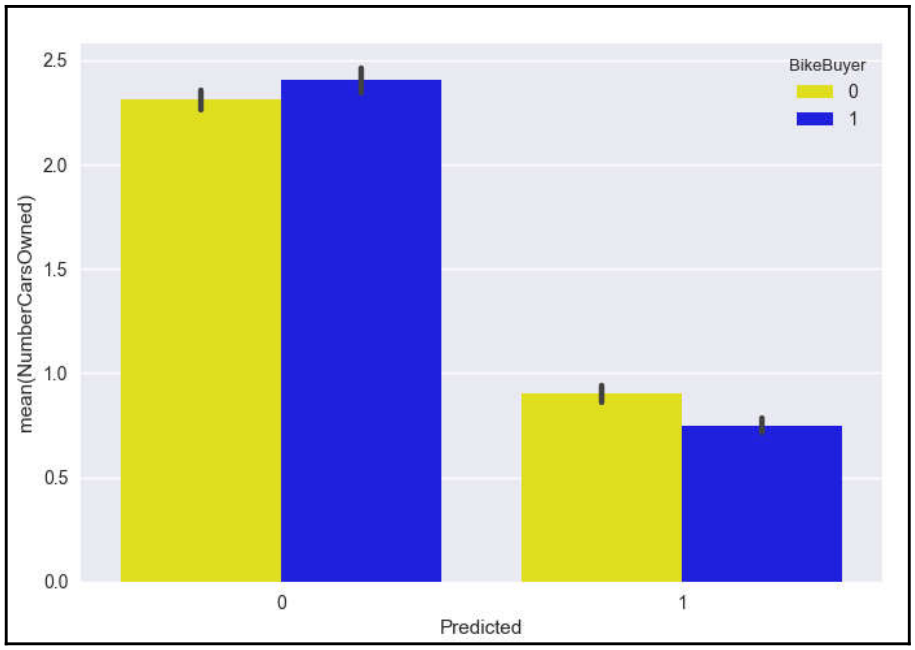

EFA Varimax
NumberChildrenAtHome
Income
NumberCarsOwned
TotalChildren
Age
HouseOwnerFlag
BikeBuyer
MR1
MR2
0.8
0.6
0.6
0.4
0.7
0.7
EFA Promax
NumberChildrenAtHome
Income
NumberCarsOwned
Age
TotalChildren
HouseOwnerFlag
BikeBuyer
MR1
MR2
0.8
0.6
0.5
0.7
0.7
0.4

Chapter 8: Supervised Machine Learning

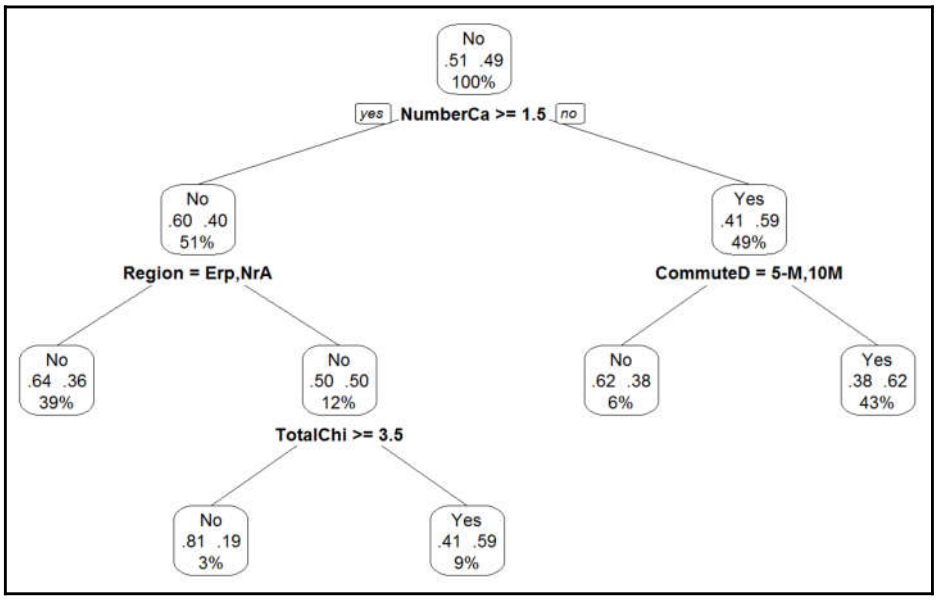


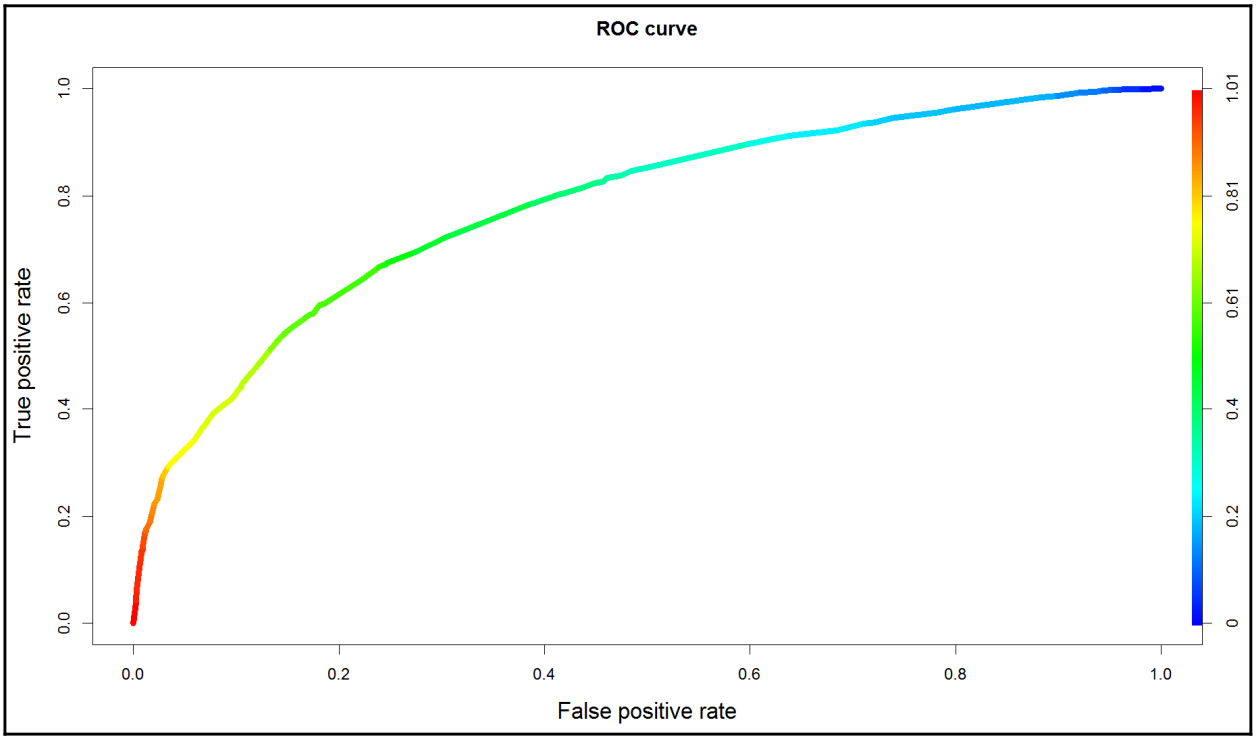






$$\boxed{S(x) = 1/(1 + e^{-x})}$$


No
.51 .49
100%
yes
NumberCa >= 1.5
no
No
.60 .40
51%
Yes
.41 .59
49%
Region = Erp,NrA
CommuteD = 5-M,10M
No
.64 .36
39%
No
.50 .50
12%
No
.62 .38
6%
Yes
.38 .62
43%
TotalChi >= 3.5
No
.81 .19
3%
Yes
.41 .59
9%


ROC curve
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0
0.2
0.4
0.61
0.81
1.01