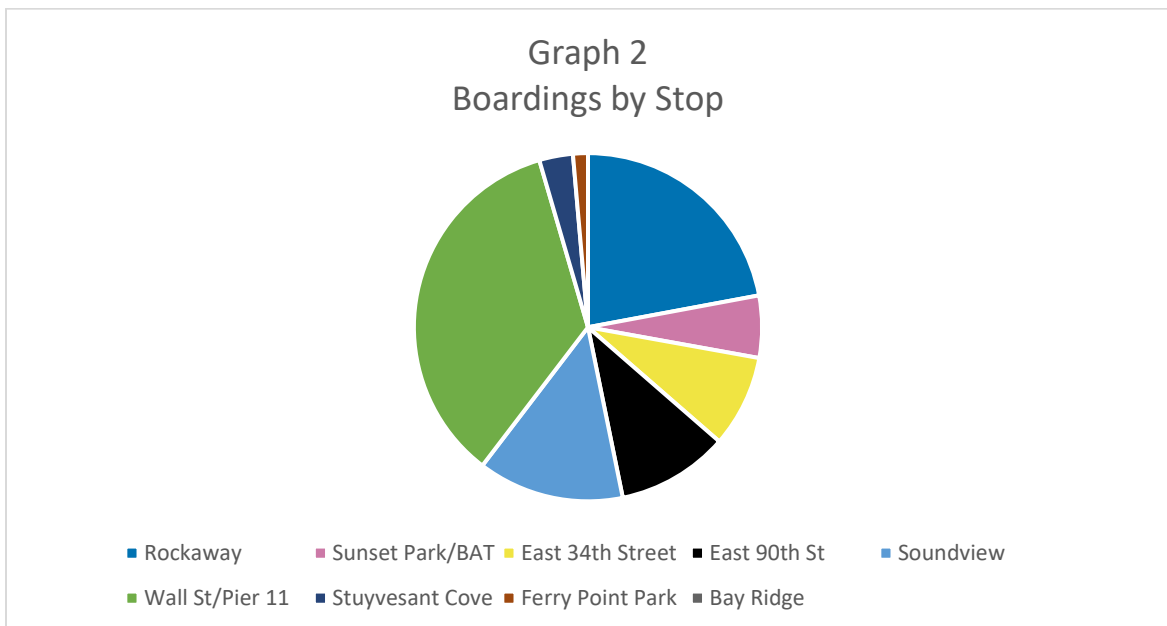
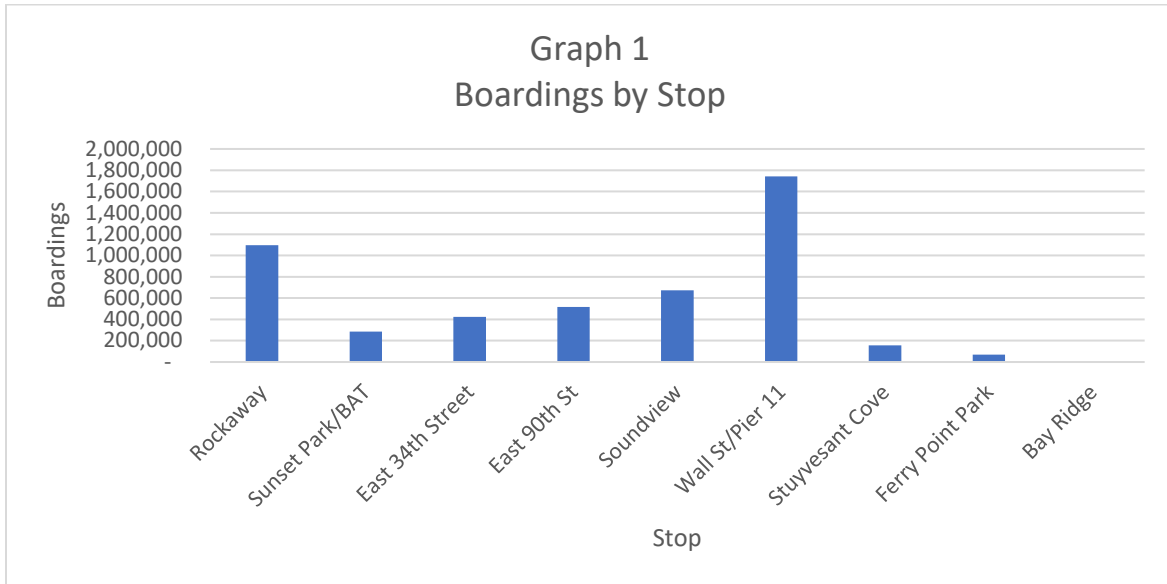
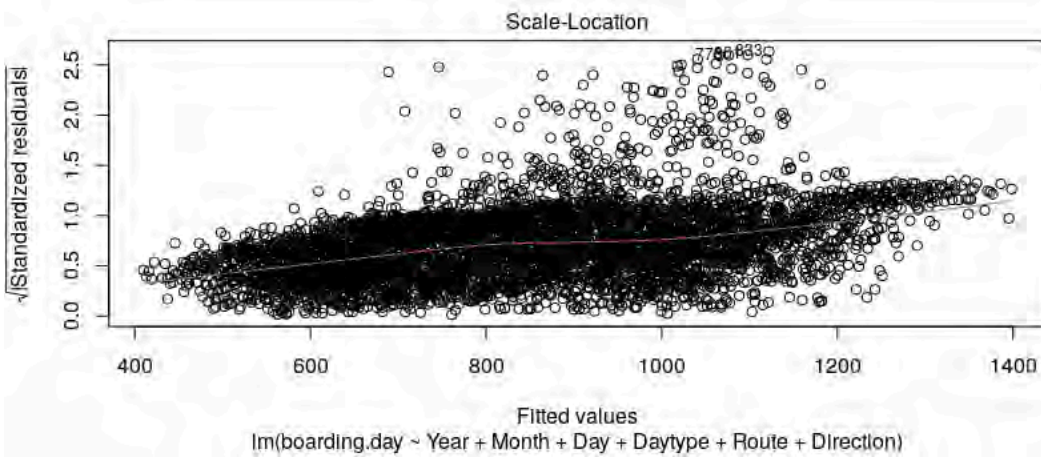
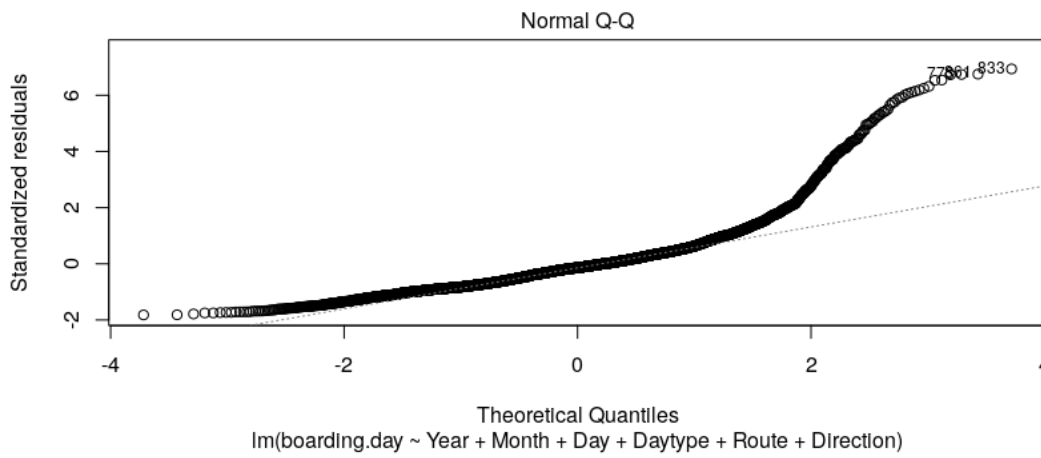
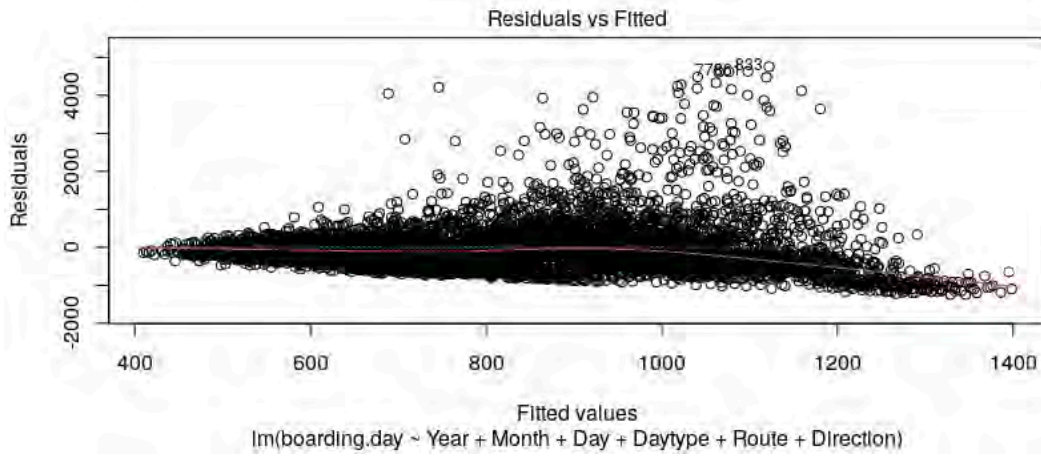


# Exam PA April 18 2023 Appendix

## Task 1 – Part a.



Task 3 – Part b.



**Task 3 – Part c.**

Model 1						Model 2					
Call: lm(formula = boarding.day ~ Year.factor + Month.factor + Day.factor + Daytype + Route + Direction, data = data.training.b)						Call: lm(formula = boarding.day ~ Year.factor + Month.factor + Daytype + Route + Direction, data = data.training.b)					
Residuals:						Residuals:					
Min	1Q	Median	3Q	Max		Min	1Q	Median	3Q	Max	
-1680.7	-293.5	-60.2	189.0	4020.9		-1671.7	-293.8	-63.6	184.1	4125.4	
Coefficients:						Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )			Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	418.171	51.719	8.085	7.73e-16 ***		(Intercept)	496.67	30.23	16.431	< 2e-16 ***	
Year.factor2020	-419.086	20.658	-20.286	< 2e-16 ***		Year.factor2020	-419.22	20.74	-20.217	< 2e-16 ***	
Year.factor2021	-243.099	20.004	-12.153	< 2e-16 ***		Year.factor2021	-242.89	20.08	-12.097	< 2e-16 ***	
Year.factor2022	-29.030	25.951	-1.119	0.263356		Year.factor2022	-29.06	26.05	-1.116	0.264682	
Month.factor2	36.817	35.303	1.043	0.297042		Month.factor2	32.13	35.33	0.909	0.363243	
Month.factor3	83.987	34.304	2.448	0.014387 *		Month.factor3	84.02	34.43	2.440	0.014716 *	
Month.factor4	149.082	34.641	4.304	1.71e-05 ***		Month.factor4	145.95	34.72	4.203	2.68e-05 ***	
Month.factor5	410.174	34.295	11.960	< 2e-16 ***		Month.factor5	410.19	34.42	11.916	< 2e-16 ***	
Month.factor6	887.750	34.637	25.630	< 2e-16 ***		Month.factor6	884.61	34.72	25.480	< 2e-16 ***	
Month.factor7	1231.148	37.499	32.832	< 2e-16 ***		Month.factor7	1231.16	37.64	32.709	< 2e-16 ***	
Month.factor8	1239.685	37.496	33.062	< 2e-16 ***		Month.factor8	1239.69	37.64	32.938	< 2e-16 ***	
Month.factor9	891.493	37.888	23.530	< 2e-16 ***		Month.factor9	888.33	37.99	23.386	< 2e-16 ***	
Month.factor10	456.844	37.499	12.183	< 2e-16 ***		Month.factor10	456.86	37.64	12.138	< 2e-16 ***	
Month.factor11	201.713	37.884	5.324	1.06e-07 ***		Month.factor11	198.52	37.98	5.227	1.79e-07 ***	
Month.factor12	67.856	42.829	1.584	0.113182		Month.factor12	67.78	42.99	1.577	0.114940	
Day.factor2	-46.610	60.016	-0.777	0.437421		DaytypeWeekend	168.52	16.75	10.063	< 2e-16 ***	
Day.factor3	-1.637	60.023	-0.027	0.978244		RouteSV	27.43	15.36	1.785	0.074266 .	
Day.factor4	132.965	60.018	2.215	0.026777 *		DirectionSB	57.27	15.36	3.728	0.000195 ***	
Day.factor5	159.163	60.016	2.652	0.008027 **		---					
Day.factor6	135.854	60.016	2.264	0.023640 *		Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Day.factor7	64.447	60.018	1.074	0.282971							
Day.factor8	6.778	60.023	0.113	0.910101							
Day.factor9	20.947	60.018	0.349	0.727098		Residual standard error: 542.1 on 4962 degrees of freedom					
Day.factor10	50.894	60.023	0.848	0.396538		Multiple R-squared: 0.4241, Adjusted R-squared: 0.4221					
Day.factor11	81.461	60.023	1.357	0.174797		F-statistic: 214.9 on 17 and 4962 DF, p-value: < 2.2e-16					
Day.factor12	45.329	60.016	0.755	0.450111							
Day.factor13	33.203	60.018	0.553	0.580142							
Day.factor14	59.941	60.018	0.999	0.317985							
Day.factor15	202.875	60.023	3.380	0.000731 ***							
Day.factor16	-22.315	60.018	-0.372	0.710051							
Day.factor17	10.762	60.016	0.179	0.857689							
Day.factor18	90.258	60.018	1.504	0.132687							
Day.factor19	77.878	60.016	1.298	0.194477							
Day.factor20	142.415	60.016	2.373	0.017685 *							
Day.factor21	69.371	60.016	1.156	0.247794							
Day.factor22	88.479	60.023	1.474	0.140526							
Day.factor23	81.935	60.018	1.365	0.172265							
Day.factor24	124.368	60.018	2.072	0.038302 *							
Day.factor25	113.395	60.016	1.889	0.058897 .							
Day.factor26	183.682	60.016	3.061	0.002221 **							
Day.factor27	114.134	60.016	1.902	0.057264 .							
Day.factor28	45.610	60.016	0.760	0.447313							
Day.factor29	30.953	61.186	0.506	0.612962							
Day.factor30	170.644	61.633	2.769	0.005649 **							
Day.factor31	173.923	70.976	2.450	0.014302 *							
DaytypeWeekend	167.812	16.706	10.045	< 2e-16 ***							
RouteSV	27.429	15.306	1.792	0.073184 .							
DirectionSB	57.272	15.306	3.742	0.000185 ***							
---											
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1											
Residual standard error: 540.1 on 4932 degrees of freedom											
Multiple R-squared: 0.4318, Adjusted R-squared: 0.4264											
F-statistic: 79.75 on 47 and 4932 DF, p-value: < 2.2e-16											

**Task 3 – Part c. (continued)**

Model 1 - MSE Training 288859.7

Model 1 - MSE Testing 435571.5

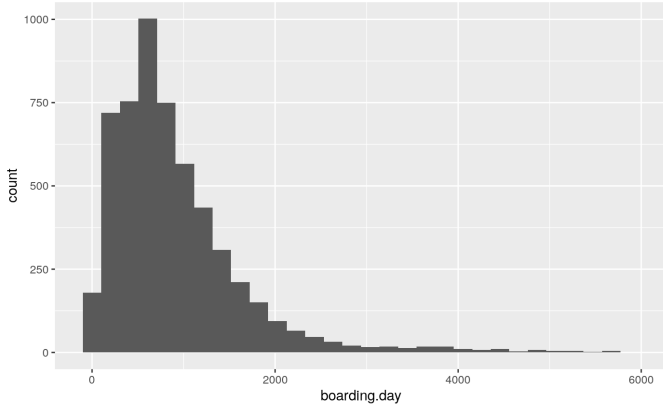
Model 2 - MSE Training 292192.5

Model 2 - MSE Testing 432221.5

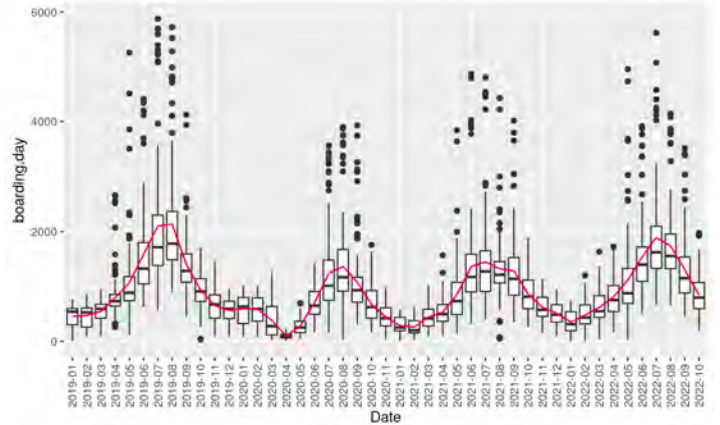
**Task 3 – Part e.**

The following plots show the distribution of Boardings per Day by Station.

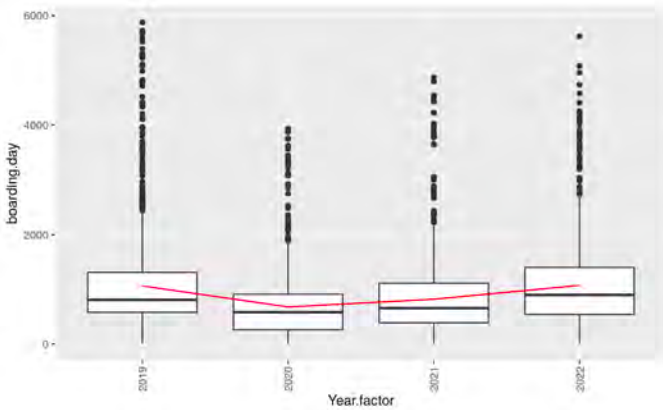
Distribution of boarding.day



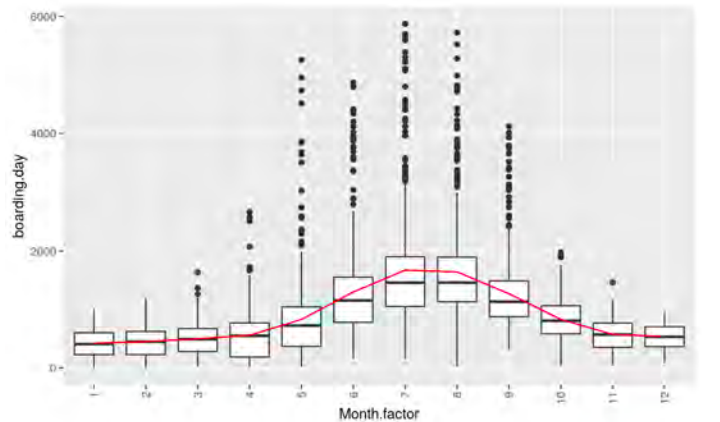
Distribution of boarding.day by Date



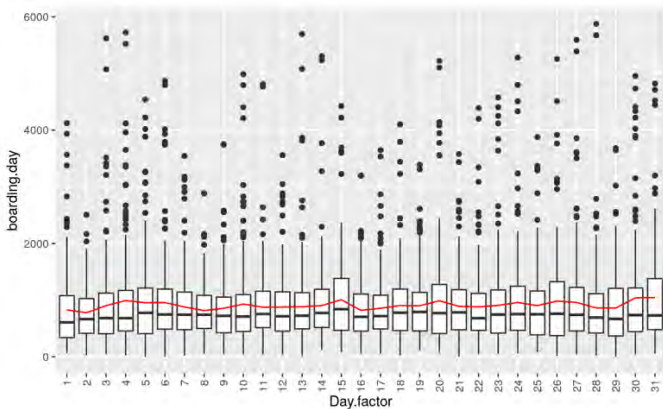
Distribution of boarding.day by Year



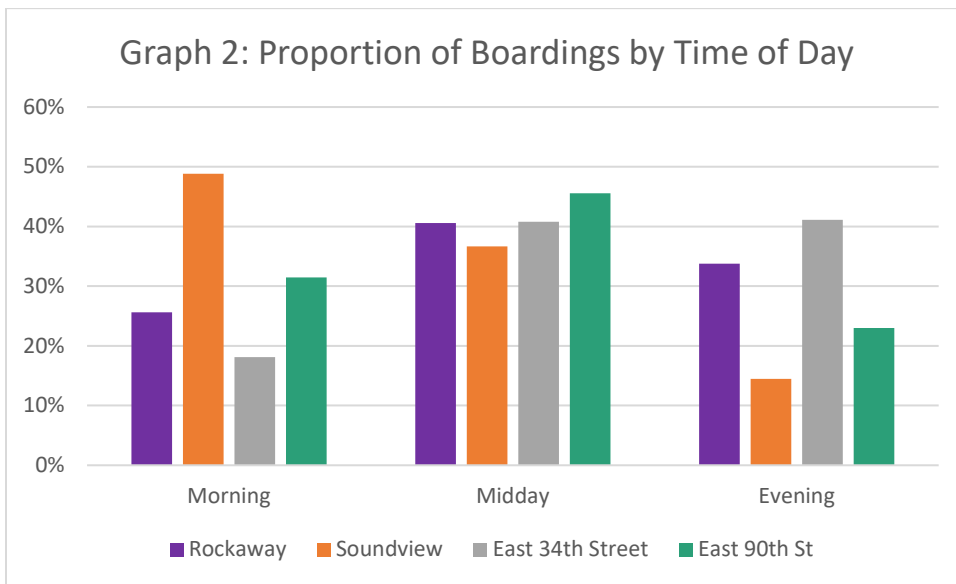
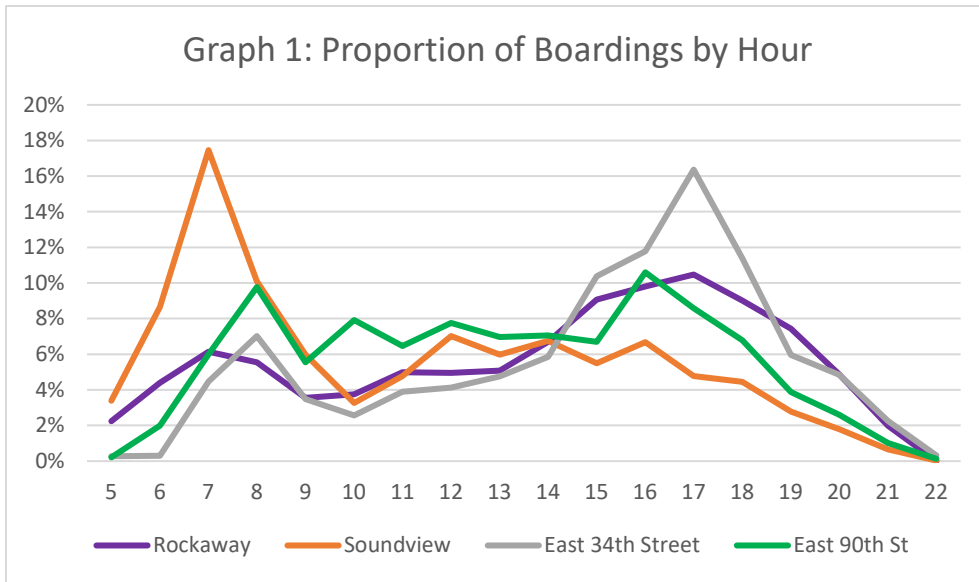
Distribution of boarding.day by Month



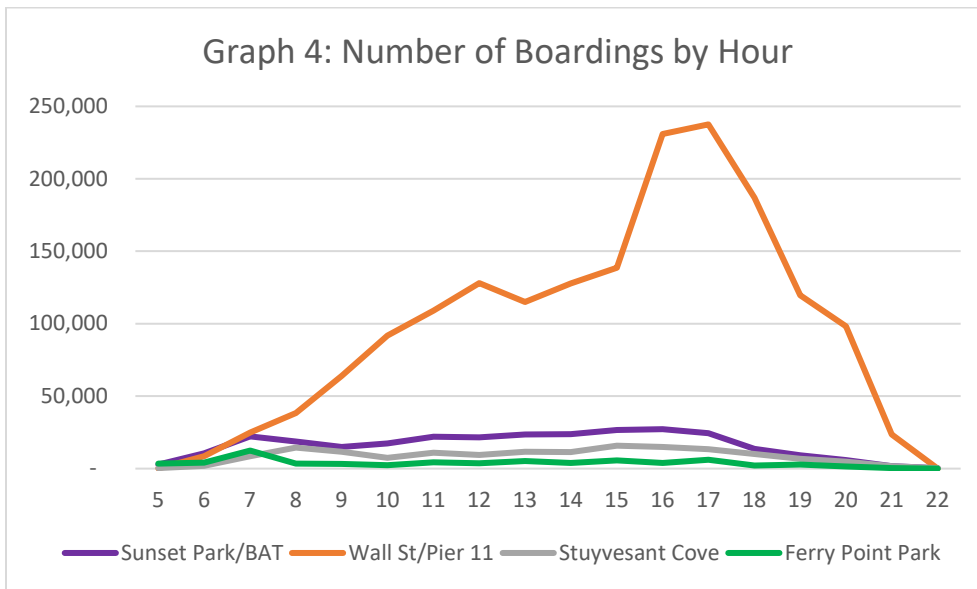
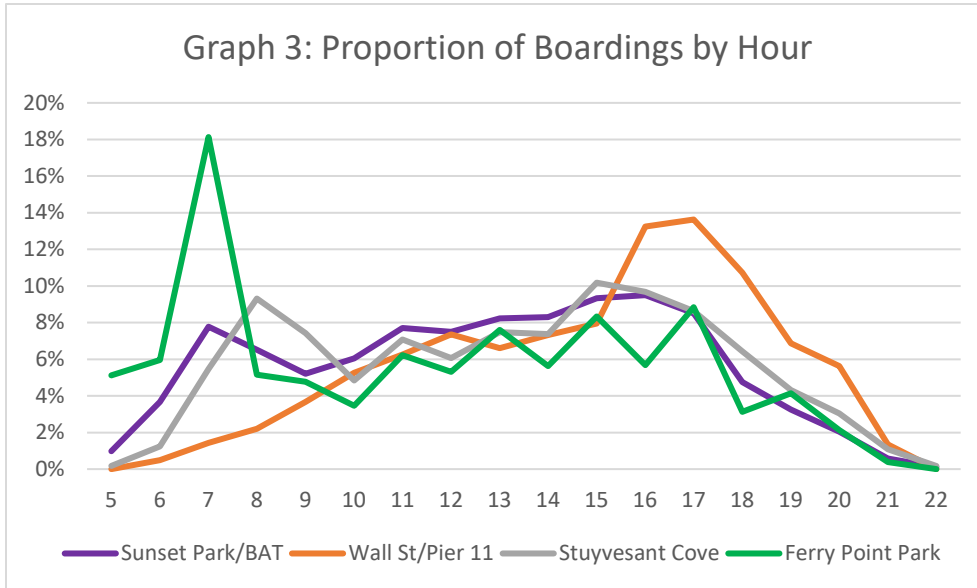
Distribution of boarding.day by Day



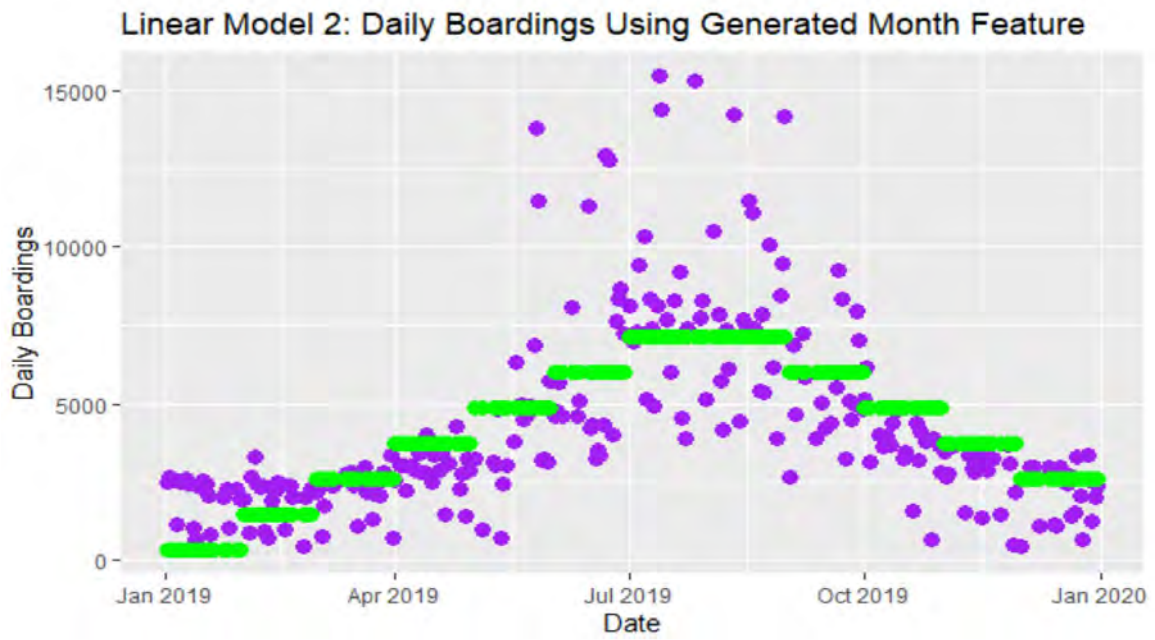
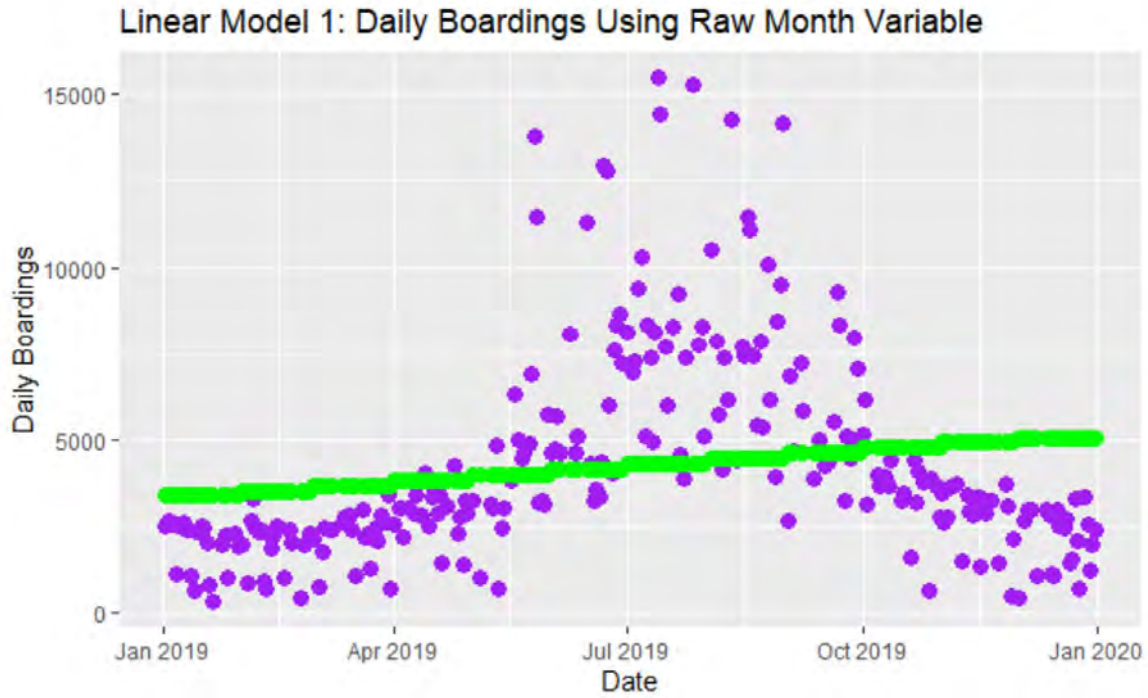
**Task 6 – Part a.**



**Task 6 – Part b.**



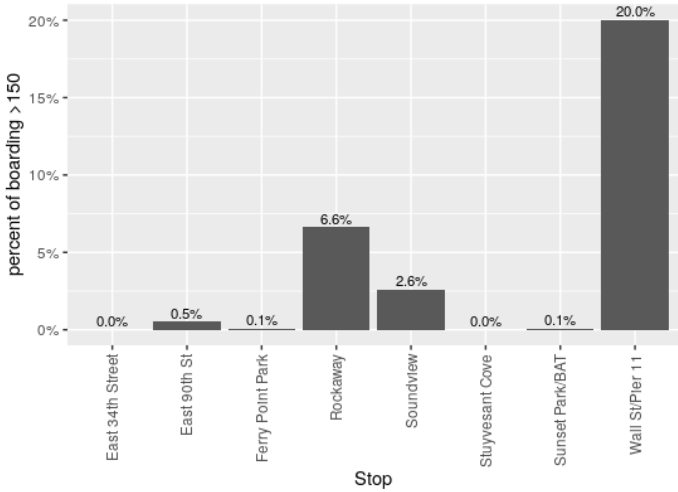
Task 7 – Part b.



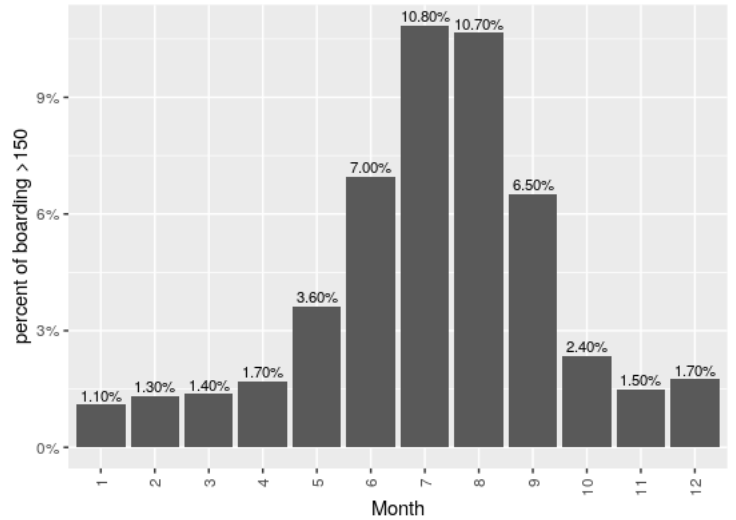


**Task 8 – Part e.**

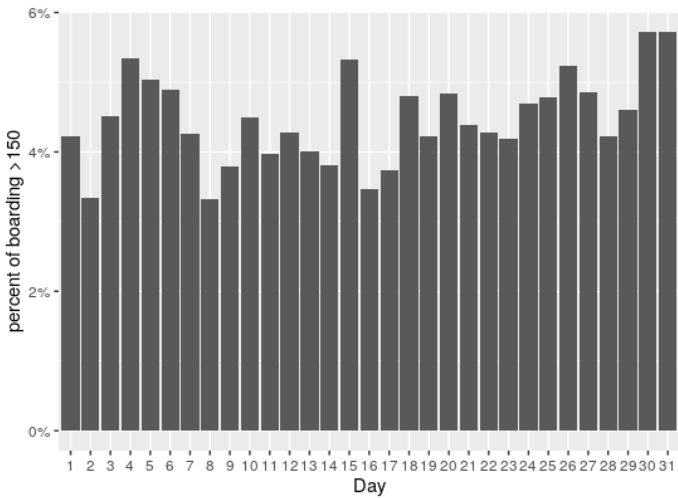
- Percent of boardings in an hour greater than 150 by Stops



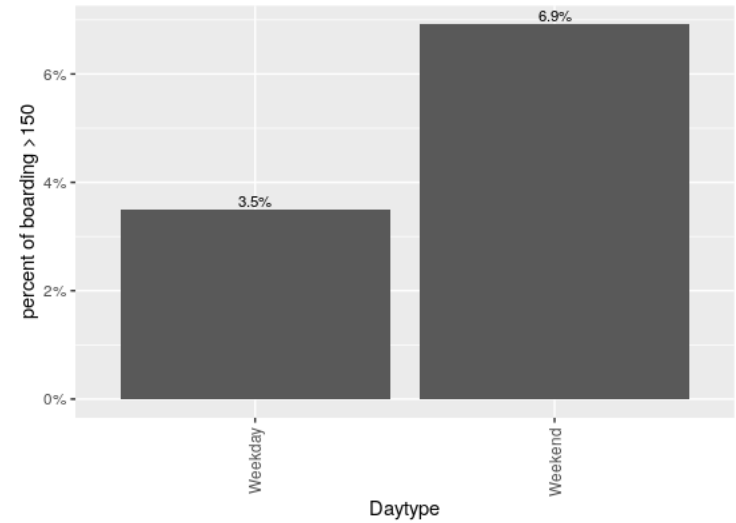
- Percent of boardings in an hour greater than 150 by Month



- Percent of boardings in an hour greater than 150 by Day



- Percent of boardings in an hour greater than 150 by Daytype



## Task 8 – Part e. (continued)

```

Model 1:

> model.glm =
glm(threshold~Day+Hour+Daytype+Stop+Month+East_34th+East_90th+Ferry_Point+Sunset_Par
k+Wall_St+Stuyvesant+Rockaway+Soundview
      ,data=train
      ,family='binomial')

> summary(model.glm)

Call:
glm(formula = threshold ~ Day + Hour + Daytype + Stop + Month +
     East_34th + East_90th + Ferry_Point + Sunset_Park + Wall_St +
     Stuyvesant + Rockaway + Soundview, family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9716  -0.1143  -0.0181   0.0000   4.5368

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.313e+01  1.212e+02  -0.191  0.84864
Day2         -5.664e-02  1.800e-01  -0.315  0.75295
Day3          2.474e-01  1.721e-01   1.438  0.15057
Day4          1.200e-02  1.729e-01   0.069  0.94468
Day5          1.228e-01  1.721e-01   0.714  0.47547
Day6          1.450e-01  1.706e-01   0.850  0.39529
Day7          7.813e-02  1.771e-01   0.441  0.65906
Day8         -3.090e-02  1.798e-01  -0.172  0.86353
Day9          4.868e-02  1.765e-01   0.276  0.78275
Day10        -2.367e-02  1.751e-01  -0.135  0.89247
Day11         1.147e-01  1.736e-01   0.660  0.50894
Day12         1.458e-01  1.729e-01   0.843  0.39917
Day13        -3.889e-02  1.774e-01  -0.219  0.82642
Day14        -1.147e-01  1.783e-01  -0.643  0.51994
Day15         2.884e-01  1.703e-01   1.694  0.09034 .
Day16        -1.622e-02  1.788e-01  -0.091  0.92770
Day17         2.163e-02  1.771e-01   0.122  0.90280
Day18         2.413e-01  1.719e-01   1.404  0.16044
Day19         9.625e-02  1.717e-01   0.560  0.57520
Day20        -3.544e-02  1.754e-01  -0.202  0.83983
Day21         1.071e-01  1.724e-01   0.621  0.53440
Day22         7.877e-02  1.747e-01   0.451  0.65209
Day23        -5.666e-02  1.765e-01  -0.321  0.74817
Day24         1.023e-01  1.720e-01   0.595  0.55182
Day25         2.192e-01  1.704e-01   1.286  0.19837
Day26         2.388e-01  1.682e-01   1.420  0.15572
Day27         5.471e-02  1.733e-01   0.316  0.75217
Day28         7.050e-02  1.752e-01   0.402  0.68741
Day29         3.064e-01  1.710e-01   1.792  0.07312 .
Day30         1.212e-01  1.738e-01   0.697  0.48567
Day31         1.245e-01  1.952e-01   0.638  0.52371
Hour8         -2.735e+00  1.729e-01 -15.815 < 2e-16 ***
Hour9         -2.701e+00  1.606e-01 -16.816 < 2e-16 ***
Hour10        -2.184e+00  1.300e-01 -16.799 < 2e-16 ***
Hour11        -2.326e+00  1.290e-01 -18.027 < 2e-16 ***
Hour12        -2.282e+00  1.247e-01 -18.310 < 2e-16 ***

```

Hour13	-2.811e+00	1.323e-01	-21.245	< 2e-16	***
Hour14	-2.551e+00	1.280e-01	-19.932	< 2e-16	***
Hour15	-1.872e+00	1.150e-01	-16.275	< 2e-16	***
Hour16	-4.841e-01	9.963e-02	-4.859	1.18e-06	***
Hour17	-7.542e-01	1.081e-01	-6.978	2.99e-12	***
Hour18	-2.333e+00	1.223e-01	-19.076	< 2e-16	***
Hour19	-3.835e+00	1.532e-01	-25.035	< 2e-16	***
Hour20	-3.840e+00	1.528e-01	-25.126	< 2e-16	***
Hour21	-5.569e+00	2.332e-01	-23.880	< 2e-16	***
DaytypeWeekend	2.458e-01	5.670e-02	4.335	1.46e-05	***
StopEast 90th St	1.543e+01	1.212e+02	0.127	0.89871	
StopFerry Point Park	1.270e+01	1.212e+02	0.105	0.91657	
StopRockaway	1.944e+01	1.212e+02	0.160	0.87256	
StopSoundview	1.784e+01	1.212e+02	0.147	0.88296	
StopStuyvesant Cove	3.014e-01	1.966e+02	0.002	0.99878	
StopSunset Park/BAT	1.316e+01	1.212e+02	0.109	0.91353	
StopWall St/Pier 11	2.154e+01	1.212e+02	0.178	0.85891	
Month2	2.057e-01	1.599e-01	1.287	0.19826	
Month3	1.599e-01	1.537e-01	1.040	0.29827	
Month4	1.537e-01	1.532e-01	1.003	0.31582	
Month5	7.318e-01	1.390e-01	5.265	1.40e-07	***
Month6	1.221e+00	1.338e-01	9.127	< 2e-16	***
Month7	1.624e+00	1.321e-01	12.294	< 2e-16	***
Month8	1.679e+00	1.317e-01	12.750	< 2e-16	***
Month9	1.233e+00	1.340e-01	9.202	< 2e-16	***
Month10	4.619e-01	1.426e-01	3.239	0.00120	**
Month11	2.198e-01	1.617e-01	1.359	0.17400	
Month12	4.756e-01	1.726e-01	2.755	0.00586	**
East_34th	9.750e-03	1.506e-03	6.473	9.62e-11	***
East_90th	1.788e-03	9.622e-04	1.858	0.06313	.
Ferry_Point	5.638e-04	1.821e-03	0.310	0.75685	
Sunset_Park	7.312e-03	1.448e-03	5.049	4.43e-07	***
Wall_St	7.945e-03	3.081e-04	25.785	< 2e-16	***
Stuyvesant	9.429e-04	1.895e-03	0.498	0.61883	
Rockaway	4.994e-03	2.813e-04	17.753	< 2e-16	***
Soundview	-4.786e-03	8.609e-04	-5.559	2.71e-08	***
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 33904 on 93490 degrees of freedom  
Residual deviance: 14970 on 93419 degrees of freedom  
AIC: 15114

Number of Fisher Scoring iterations: 19

**Model 2:**

```
> model.backward = step(model.glm, direction='backward', scope=formula(model.glm),
trace=0)
> summary(backward)
```

Call:

```
glm(formula = threshold ~ Hour + Daytype + Stop + Month + East_34th +
East_90th + Sunset_Park + Wall_St + Rockaway + Soundview,
family = "binomial", data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.9837	-0.1147	-0.0182	0.0000	4.5351

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.305e+01	1.212e+02	-0.190	0.849146
Hour8	-2.728e+00	1.728e-01	-15.789	< 2e-16 ***
Hour9	-2.696e+00	1.603e-01	-16.820	< 2e-16 ***
Hour10	-2.174e+00	1.292e-01	-16.830	< 2e-16 ***
Hour11	-2.310e+00	1.278e-01	-18.076	< 2e-16 ***
Hour12	-2.276e+00	1.241e-01	-18.332	< 2e-16 ***
Hour13	-2.792e+00	1.308e-01	-21.340	< 2e-16 ***
Hour14	-2.547e+00	1.275e-01	-19.976	< 2e-16 ***
Hour15	-1.860e+00	1.141e-01	-16.307	< 2e-16 ***
Hour16	-4.813e-01	9.863e-02	-4.880	1.06e-06 ***
Hour17	-7.453e-01	1.064e-01	-7.003	2.50e-12 ***
Hour18	-2.327e+00	1.207e-01	-19.274	< 2e-16 ***
Hour19	-3.824e+00	1.522e-01	-25.130	< 2e-16 ***
Hour20	-3.834e+00	1.517e-01	-25.277	< 2e-16 ***
Hour21	-5.560e+00	2.334e-01	-23.818	< 2e-16 ***
DaytypeWeekend	2.438e-01	5.618e-02	4.340	1.42e-05 ***
StopEast 90th St	1.543e+01	1.212e+02	0.127	0.898694
StopFerry Point Park	1.274e+01	1.212e+02	0.105	0.916313
StopRockaway	1.945e+01	1.212e+02	0.160	0.872495
StopSoundview	1.785e+01	1.212e+02	0.147	0.882906
StopStuyvesant Cove	3.463e-01	1.970e+02	0.002	0.998597
StopSunset Park/BAT	1.316e+01	1.212e+02	0.109	0.913512
StopWall St/Pier 11	2.155e+01	1.212e+02	0.178	0.858898
Month2	2.009e-01	1.595e-01	1.259	0.207951
Month3	1.627e-01	1.535e-01	1.059	0.289402
Month4	1.522e-01	1.529e-01	0.995	0.319763
Month5	7.423e-01	1.384e-01	5.364	8.13e-08 ***
Month6	1.221e+00	1.331e-01	9.178	< 2e-16 ***
Month7	1.621e+00	1.316e-01	12.324	< 2e-16 ***
Month8	1.676e+00	1.312e-01	12.770	< 2e-16 ***
Month9	1.239e+00	1.330e-01	9.313	< 2e-16 ***
Month10	4.680e-01	1.420e-01	3.296	0.000981 ***
Month11	2.202e-01	1.613e-01	1.365	0.172114
Month12	4.714e-01	1.724e-01	2.734	0.006265 **
East_34th	9.694e-03	1.501e-03	6.460	1.05e-10 ***
East_90th	1.884e-03	9.482e-04	1.987	0.046906 *
Sunset_Park	7.327e-03	1.435e-03	5.104	3.32e-07 ***
Wall_St	7.964e-03	3.061e-04	26.020	< 2e-16 ***
Rockaway	5.011e-03	2.778e-04	18.039	< 2e-16 ***
Soundview	-4.694e-03	8.588e-04	-5.466	4.60e-08 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 33904 on 93490 degrees of freedom  
 Residual deviance: 14996 on 93451 degrees of freedom  
 AIC: 15076

Number of Fisher Scoring iterations: 19

```
> backward$anova
      Step Df   Deviance Resid. Df Resid. Dev    AIC
```

1		NA	NA	93419	14970.39	15114.39
2	- Day	30	25.45471735	93449	14995.84	15079.84
3	- Ferry_Point	1	0.06772975	93450	14995.91	15077.91
4	- Stuyvesant	1	0.55082588	93451	14996.46	15076.46

```

Model 3:

> lambda = 0.0006
> model.lasso <- glmnet(x_train,y=train$threshold, family='binomial'
                        ,alpha = 1,lambda = lambda)
> coef(model.lasso)

74 x 1 sparse Matrix of class "dgCMatrix"
              s0
(Intercept)      .
Day1              .
Day2             -0.0577615876
Day3              .
Day4              .
Day5              .
Day6              .
Day7              .
Day8             -0.0373050308
Day9              .
Day10            -0.0551951726
Day11             .
Day12             .
Day13            -0.0658484023
Day14            -0.1112352096
Day15             .
Day16            -0.0163006574
Day17             .
Day18             .
Day19             .
Day20            -0.0217629745
Day21             .
Day22             .
Day23            -0.0655785839
Day24             .
Day25             .
Day26             .
Day27             .
Day28             .
Day29             .
Day30             .
Day31             .
Hour8            -2.0628863397
Hour9            -2.0169346619
Hour10           -1.4231334633
Hour11           -1.5455933137
Hour12           -1.4714125221
Hour13           -1.8702505302
Hour14           -1.7618588383
Hour15           -1.0980904308
Hour16            .
Hour17           -0.0370425166
Hour18           -1.3117559690
Hour19           -2.3844868411

```

Hour20	-2.5017065780
Hour21	-3.8344403716
DaytypeWeekend	0.0075618274
StopEast 34th Street	-7.5268198435
StopEast 90th St	-5.9607620867
StopFerry Point Park	-6.6476355907
StopRockaway	-2.7901251325
StopSoundview	-4.1066400841
StopStuyvesant Cove	-7.1264237688
StopSunset Park/BAT	-7.1122932457
StopWall St/Pier 11	-0.9809377177
Month2	-0.6310149303
Month3	-0.6692614247
Month4	-0.6568743827
Month5	-0.1299453648
Month6	0.1958325874
Month7	0.5522111037
Month8	0.6165329075
Month9	0.2310853377
Month10	-0.3551744027
Month11	-0.5723464622
Month12	-0.3497759749
East_34th	0.0003451117
East_90th	0.0011171319
Ferry_Point	.
Sunset_Park	0.0062245602
Wall_St	0.0069500770
Stuyvesant	.
Rockaway	0.0039400031
Soundview	-0.0014923522

- **AUC Table**

<b>model</b>	<b>auc.train</b>	<b>auc.test</b>
GLM	0.9648982	0.9619476
Backward Selection	0.9643398	0.9620257
LASSO	0.9610321	0.9589683