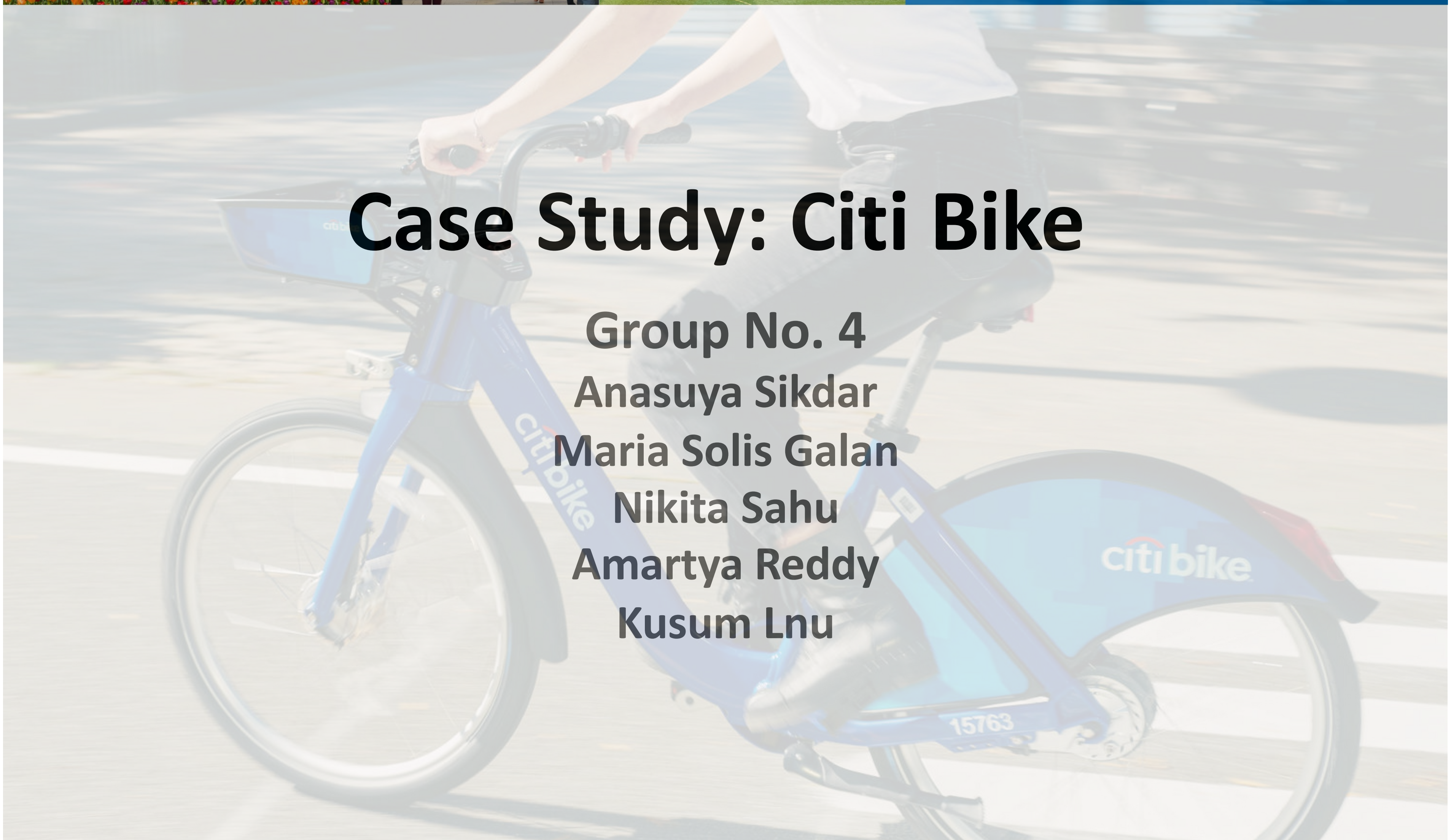
DePaul University



Case Study: Citi Bike

Group No. 4

Anasuya Sikdar

Maria Solis Galan

Nikita Sahu

Amartya Reddy

Kusum Lnu

Case Overview

What is Citi Bike?

- *Bike share system*
- *Offers: single rides, day pass, and memberships (NYC only)*
- ***25,000 bikes** in over **1,500 stations** located across:*
  - *Manhattan, Brooklyn, Queens, the Bronx, Jersey City, Hoboken*

Problem? ***Imbalanced stations > revenue loss, high logistic costs***

Data: NYC Citi Bike

- ***Objective:***

➢ Determine the ***optimal number of bikes*** to allocate to each station at the beginning of the day to ***maximize the number of daily trips***

- ***Questions:***

➢ Each day (during Morning and Evening), how many trips are made between the stations?

➢ Is the demand constant throughout the year or which factors affect it?

➢ How many bikes should we deploy at each selected station to maximize number of trips?

NY Citi Bike Data

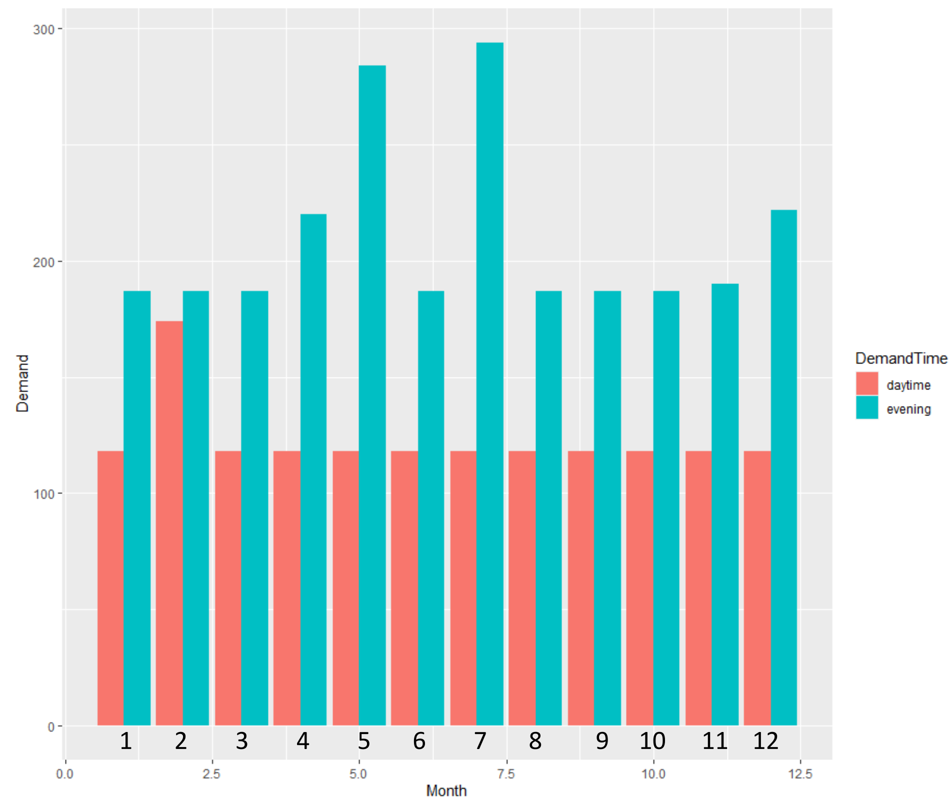
- Overall: 31,414 observation with 25 variables
- Data considers demand based on day of week, month, income, weather factors, location, distance.
- June 1st, 2017 to May 31st, 2018
- Demand fluctuates depending on the month and day of week.
- "Demand" is the Response.

	Frequency	Percent
	1	0.0
daytime	12139	38.6
evening	19275	61.4
Total	31415	100.0

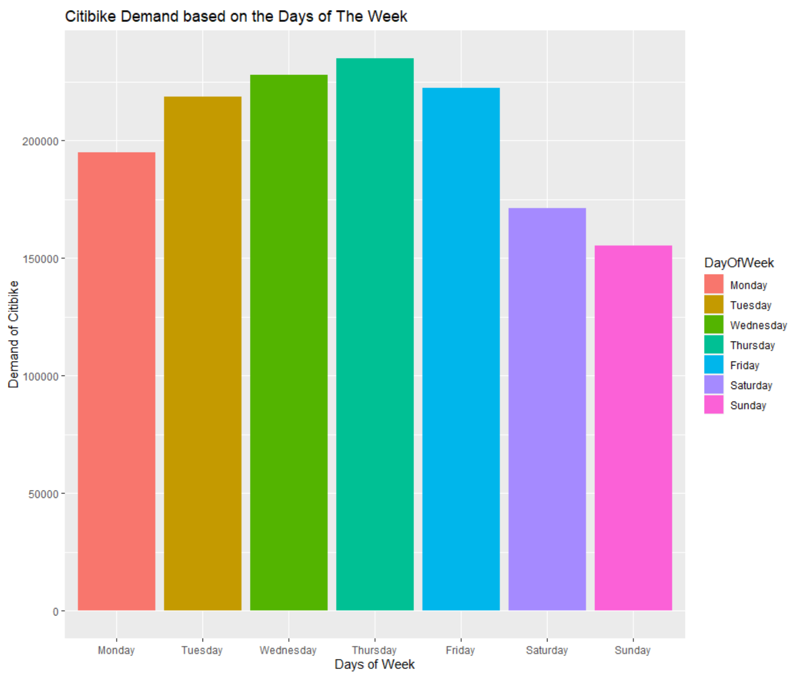
citibikeDemand$Month :

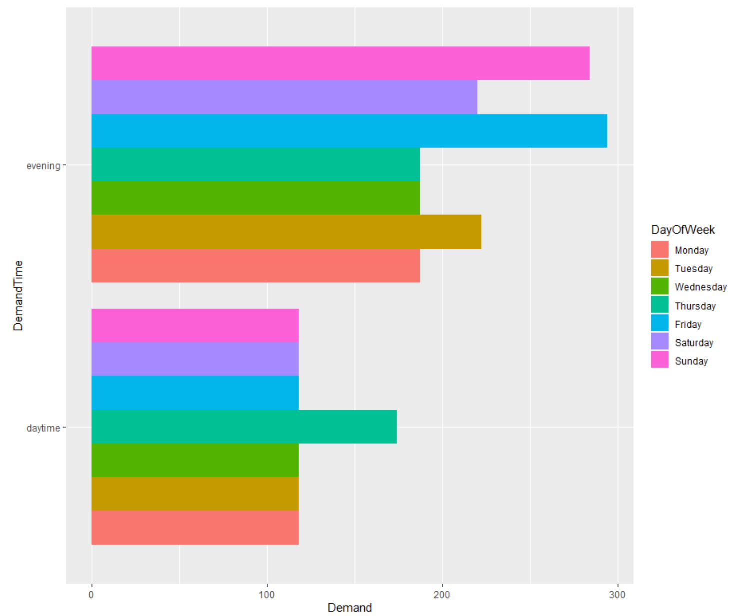
	Frequency	Percent	Cum. percent
1	1372	4.4	4.4
2	1535	4.9	9.3
3	1891	6.0	15.3
4	2502	8.0	23.2
5	3554	11.3	34.6
6	3051	9.7	44.3
7	3201	10.2	54.5
8	3300	10.5	65.0
9	3408	10.8	75.8
10	3492	11.1	86.9
11	2452	7.8	94.7
12	1656	5.3	100.0
Total	31414	100.0	100.0

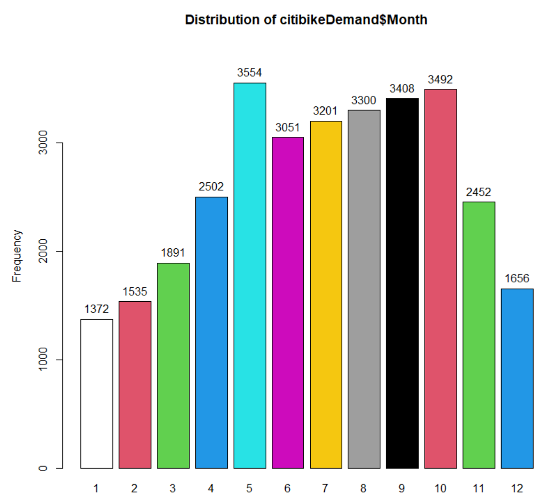
>



Demand distribution based on days of the week, time, and month





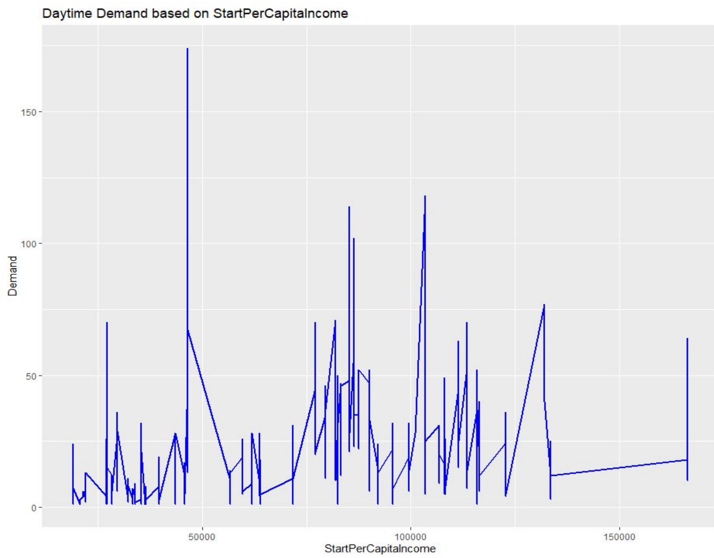


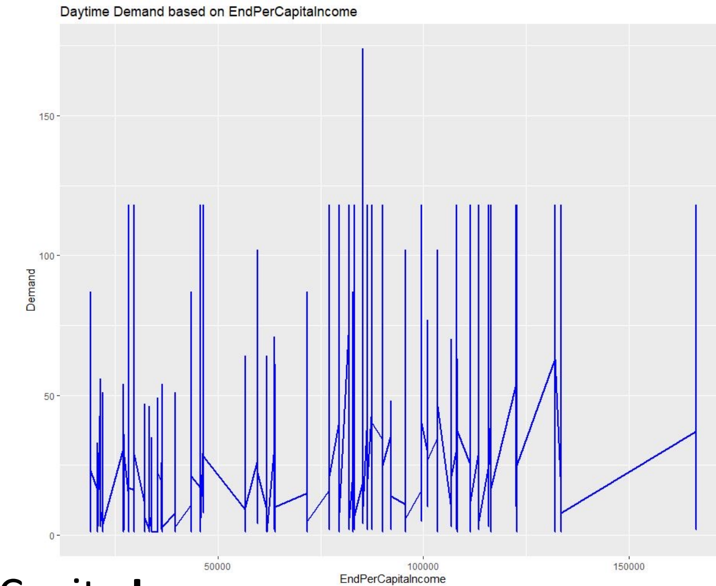
Overall highest demand
Week:
Thursday daytime
Wednesday evening
Friday evening

Overall highest demand
Month:
May – 3554
October – 3492
September - 3408

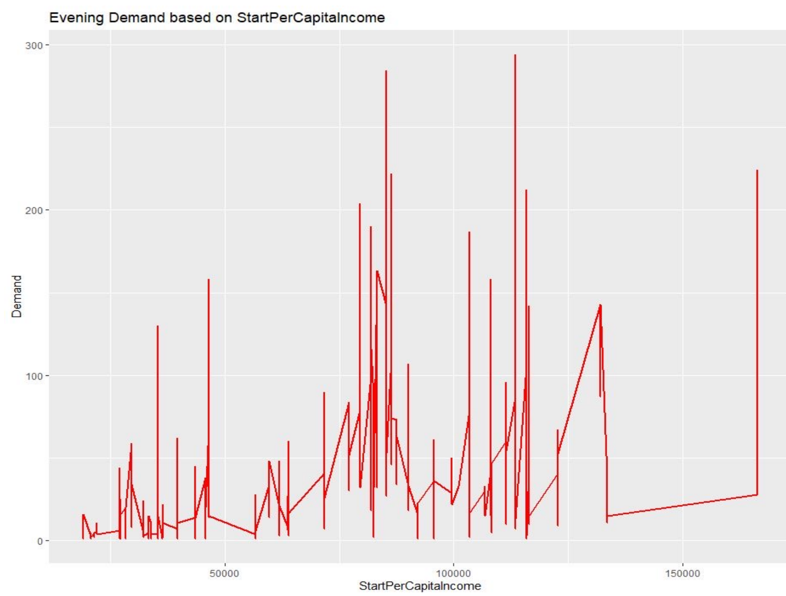
Descriptive Analysis and Visualization

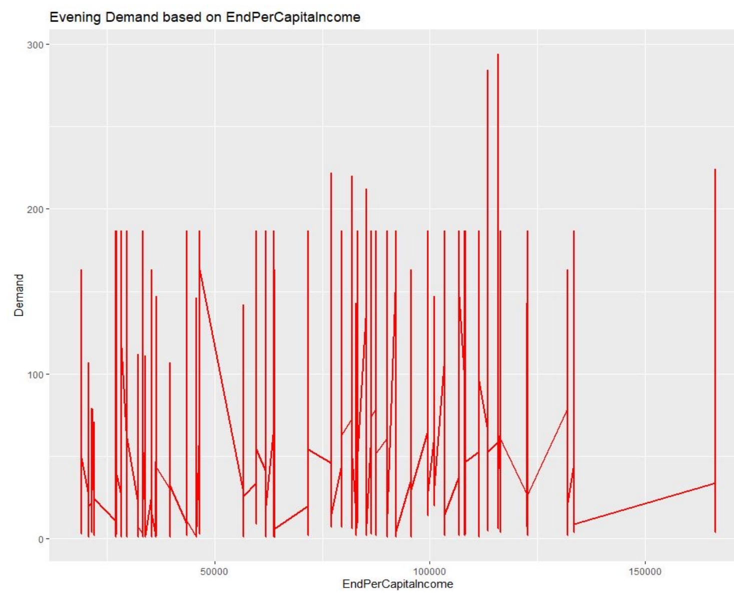
Daytime and Evening Demand based on **Start** Per Capita **Income**



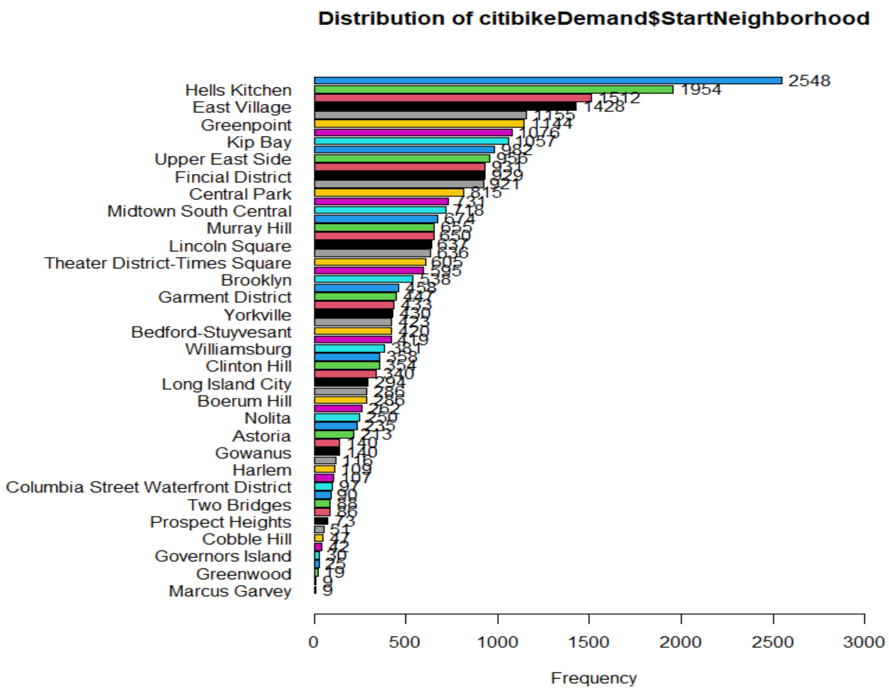


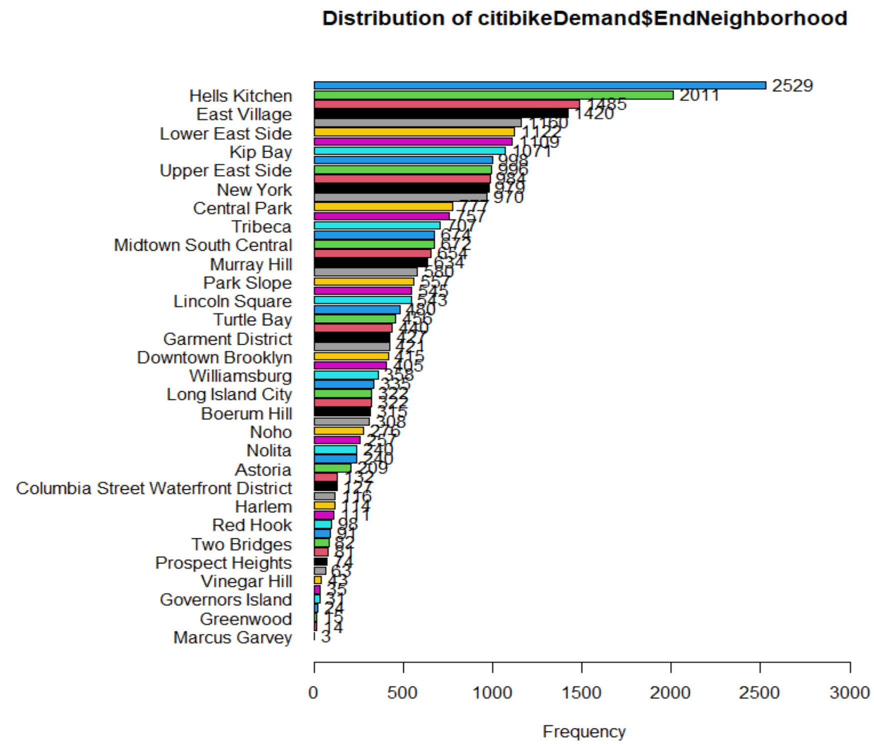
Daytime and Evening Demand based on **End** Per Capita **Income**





Demand based on Start and End Neighborhood





Overall *highest* demand found in:

Hells Kitchen

East Village

Kip Bay

Upper East Side

Central Park

Predictive Analysis

R Code of *Multiple Linear regression Model*

```
#Advanced Predictive Model 3 Final Model
reg_fit3 <- lm(Demand ~ factor(DemandTime) + factor(StartStationId) + factor(EndStationId) + factor(DayOfWeek) + factor(Month) , data = citibikeDemand)

summary(reg_fit3)
```

Mathematical Representation of
Multiple Linear Regression Model

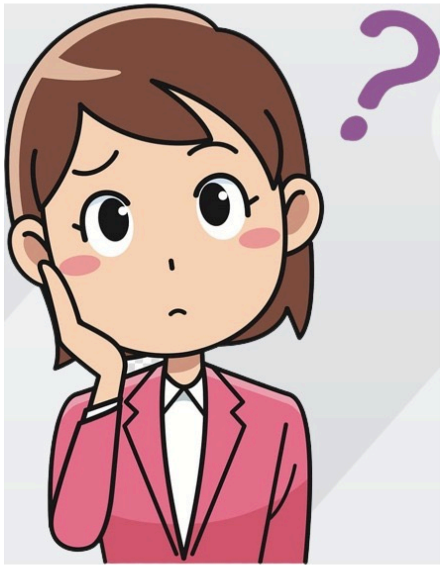
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

```
> summary(reg_fit3)

Call:
lm(formula = Demand ~ factor(DemandTime) + factor(StartStationId) +
    factor(EndStationId) + factor(DayOfWeek) + factor(Month),
    data = citibikeDemand)

Residuals:
    Min      1Q  Median      3Q     Max
-86.699  -4.843   1.118   6.354 163.148

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                25.49875    1.87300  13.614  < 2e-16 ***
factor(DemandTime)evening  19.56799    0.13528 144.653  < 2e-16 ***
```

Model Goodness of Fit

Quantity	Value
RSE	10.82
R-Squared	0.8952
F-statistic	169.5
P-value	<2.2e-16

```
Residual standard error: 10.82 on 29906 degrees of freedom
Multiple R-squared:  0.8952,	Adjusted R-squared:  0.8899
F-statistic: 169.5 on 1507 and 29906 DF,  p-value: < 2.2e-16
```



Variance Inflation Factor

-used to check for Multicollinearity Problem in the model

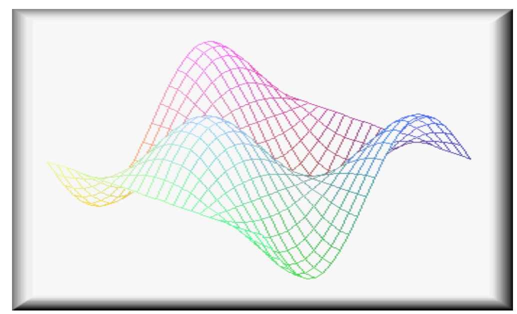
- If vif > 5 or 10 denote model has problems estimating the coefficient

```
vif(reg_fit3)
```

```
##                              GVIF  Df GVIF^(1/(2*Df))
## factor(DemandTime)     1.164510e+00   1        1.079125
## factor(StartStationId) 4.343027e+24 743        1.038915
## factor(EndStationId)   4.465766e+24 746        1.038775
## factor(DayOfWeek)      1.438157e+00   6        1.030743
## factor(Month)          1.995626e+00  11        1.031906
```

What do we Infer from the Table?

- value of vif between 1-5, moderate correlation (good value)



Prescriptive Analysis

# bikes	2000

Morning Demand

		Destination Station					
		E 7 St & Avenue A	E 24 St & Park Ave S	12 Ave & W 40 St	Central Park S & 6 Ave	W 63 St & Broadway	total
Origin Station Name	E 7 St & Avenue A	72	70	73	71	70	356
	E 24 St & Park Ave S	57	56	59	57	56	284
	12 Ave & W 40 St	77	75	79	76	75	383
	Central Park S & 6 Ave	68	67	70	68	67	339
	W 63 St & Broadway	43	41	45	43	42	213
	total	317	309	325	315	310	1575

Evening Demand

		Destination Station					
		E 7 St & Avenue A	E 24 St & Park Ave S	12 Ave & W 40 St	Central Park S & 6 Ave	W 63 St & Broadway	total
Origin Station Name	E 7 St & Avenue A	91	90	93	91	90	454
	E 24 St & Park Ave S	77	75	78	76	75	382
	12 Ave & W 40 St	96	95	98	96	95	480
	Central Park S & 6 Ave	88	86	85	87	86	433
	W 63 St & Broadway	63	61	60	62	61	307
	total	415	407	414	412	408	2056

Selected Stations

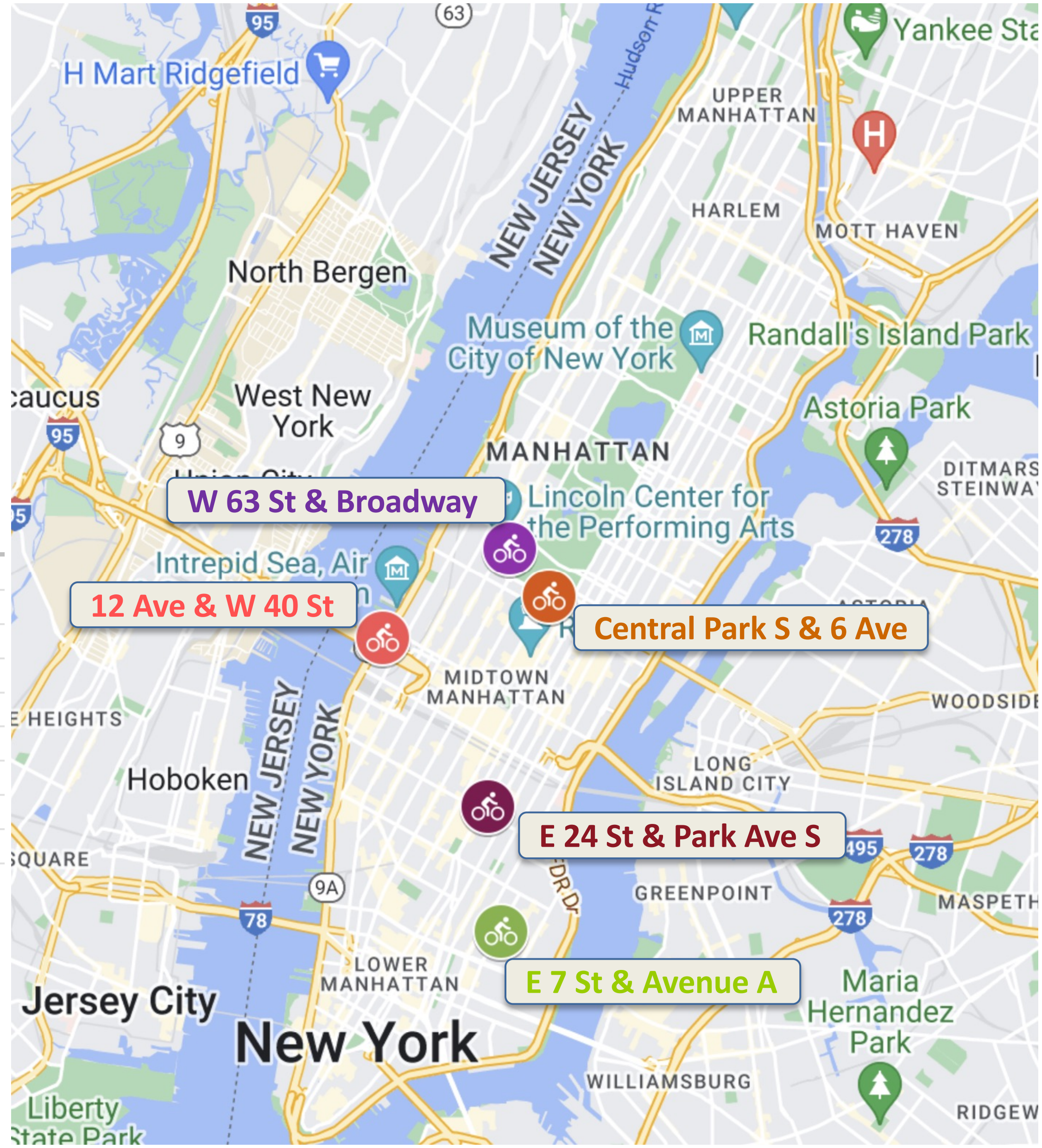
- ❑ E 7 St & Avenue A > East Village
- ❑ E 24 St & Park Ave S > Midtown South Central
- ❑ 12 Ave & W 40 St > Hells Kitchen
- ❑ Central Park S & 6 Ave > Central Park
- ❑ W 63 St & Broadway > Lincoln Square

Selected stations and allocation for each

- -E 7 St & Avenue A - 493
- -E 24 St & Park Ave S – 357
- -12 Ave & W 40 St - 538
- -Central Park S & 6 Ave – 399
- -W 63 St & Broadway - 213

	Number trips	3572
Decisions	Station	*Initial Allocation*
	E 7 St & Avenue A	493
	E 24 St & Park Ave S	357
	12 Ave & W 40 St	538
	Central Park S & 6 Ave	399
	W 63 St & Broadway	213
	total	2000

Max Number of Trips: 3572



Recommendation

Station ID	Station Name	Neighborhood	Initial Allocation
432	E 7 St & Avenue A	East Village	493
491	E 24 St & Park Ave S	Midtown South Central	357
514	12 Ave & W 40 St	Hells Kitchen	538
2006	Central Park S & 6 Ave	Central Park	399
3158	W 63 St & Broadway	Lincoln Square	213

✓Allocation of **2000** bikes in order to satisfy **3572** trips between these stations.

- ✓ Overall bike allocation:
  - ✓ Consider demand levels in each neighborhood and assign trucks to restock bikes based on neighborhood demand.
  - ✓ Set number of bikes to be allocated to each neighborhood and adjust on Monthly basis according to Demand fluctuations.