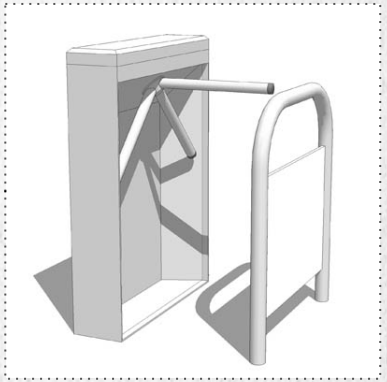
Data & Analysis

Is there a correlation between **subway rides** and number of Citi Bike pick-ups?

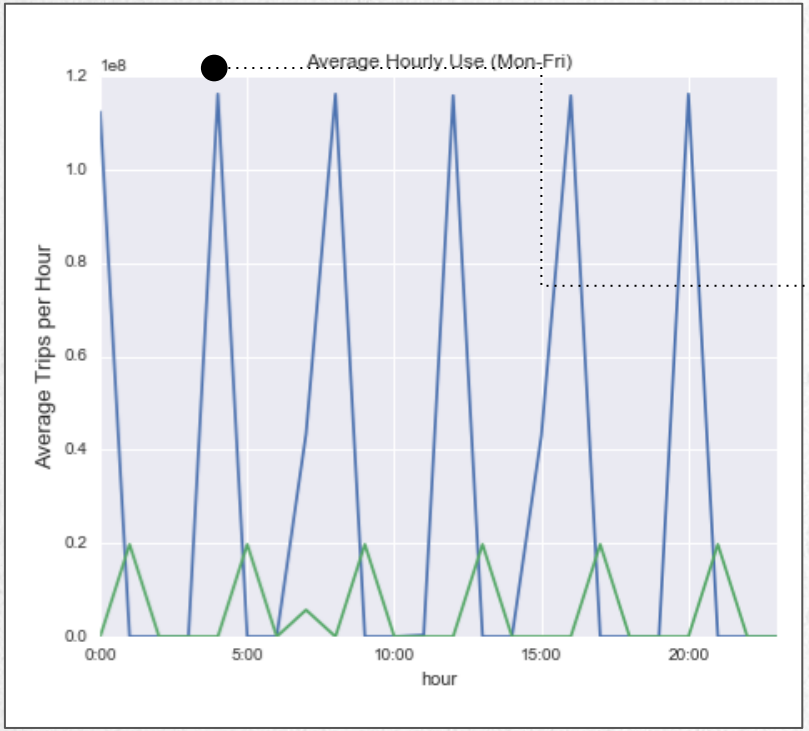
MTA turnstile data

Citi Bike trips



Data & Analysis

Average number of Exits at Times Square and West 4 stations - by hour



STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
59 ST	NQR456	BMT	08/27/2016	00:00:00	REGULAR	5799442	1966041
59 ST	NQR456	BMT	08/27/2016	04:00:00	REGULAR	5799463	1966044
59 ST	NQR456	BMT	08/27/2016	08:00:00	REGULAR	5799492	1966079
59 ST	NQR456	BMT	08/27/2016	12:00:00	REGULAR	5799610	1966155
59 ST	NQR456	BMT	08/27/2016	16:00:00	REGULAR	5799833	1966214
59 ST	NQR456	BMT	08/27/2016	20:00:00	REGULAR	5800121	1966271
59 ST	NQR456	BMT	08/28/2016	00:00:00	REGULAR	5800252	1966295
59 ST	NQR456	BMT	08/28/2016	04:00:00	REGULAR	5800281	1966303

Data & Analysis

Examine **neighborhood** influence on Citi Bike use

	ridership	StationID	BoroName	NTACode	NTAName	all_househ	child_per_	Total_Popu
start station id								
79	3076.0	3115.0	Brooklyn	BK76	Greenpoint	2635.0	292.777778	34719.0
83	1696.0	3070.0	Brooklyn	BK90	East Williamsburg	3184.0	636.800000	34158.0
116	5174.0	3235.0	Manhattan	MN20	Murray Hill-Kips Bay	2794.0	698.500000	50742.0
120	672.0	3064.0	Brooklyn	BK35	Stuyvesant Heights	11331.0	1030.090909	63504.0
128	6952.0	3121.0	Queens	QN31	Hunters Point-Sunnyside-West Maspeth	6375.0	910.714286	63271.0
137	2727.0	402.0	Manhattan	MN13	Hudson Yards-Chelsea-Flatiron-Union Square	4785.0	683.571429	70150.0
143	2062.0	466.0	Manhattan	MN13	Hudson Yards-Chelsea-Flatiron-Union Square	4785.0	683.571429	70150.0

- Spatial analysis - calculate ridership in each station by their belonging Neighborhood Tabulation Area (NTA).
- Join other indicators: household units, children and total population of that neighborhood.

Data & Analysis

Examine **neighborhood** influence on Citi Bike use

- Training set and Validation set
- R-squared is not significant

OLS Regression Results

Dep. Variable:	ridership	R-squared:	0.124
Model:	OLS	Adj. R-squared:	0.067
Method:	Least Squares	F-statistic:	2.179
Date:	Mon, 12 Dec 2016	Prob (F-statistic):	0.103
Time:	06:34:17	Log-Likelihood:	-456.41
No. Observations:	50	AIC:	920.8
Df Residuals:	46	BIC:	928.5
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
Intercept	2951.3685	982.600	3.004	0.004	973.497	4929.240
Total_Popu	0.0615	0.025	2.483	0.017	0.012	0.111
all_househ	-0.1551	0.314	-0.494	0.624	-0.788	0.477
child_per_	-2.0521	2.052	-1.000	0.323	-6.183	2.079

Omnibus:	3.478	Durbin-Watson:	1.813
Prob(Omnibus):	0.176	Jarque-Bera (JB):	3.287
Skew:	0.573	Prob(JB):	0.193
Kurtosis:	2.486	Cond. No.	1.39e+05

Next Steps

- Analyse **delayed drop of bikes** and it's impacts & consideration in the tweaking the optimization models.
- Cluster the bike stations by **high usage** in addition to their proximity to the public transit subway station entrance.
- Identify **other potential influential factors** for the Multivariate Regression Model
- Scale these analyses for **other cities** operating Bike Sharing Models.

Recommendations

➢ A stronger and uniform data collection strategy between the operators. Open data standards -

General Bikeshare Feed Specification (GBFS)

North America Bike Share Adopt Open Data Std. (NABSA)

Bike Sharing Systems to have ready data availability for comparison between operators.

➢ Sharing of Best Practices and Optimization solutions between Bike Sharing Operators

Thank you.