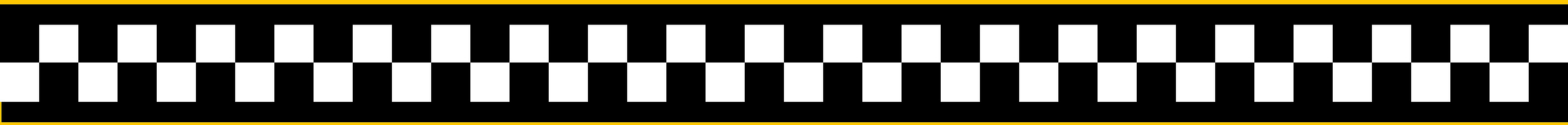
SEPTEMBER 25, 2020

YOUR ETA IS ...

NYC TAXI DATASET

Proposal by Daniel Wallace



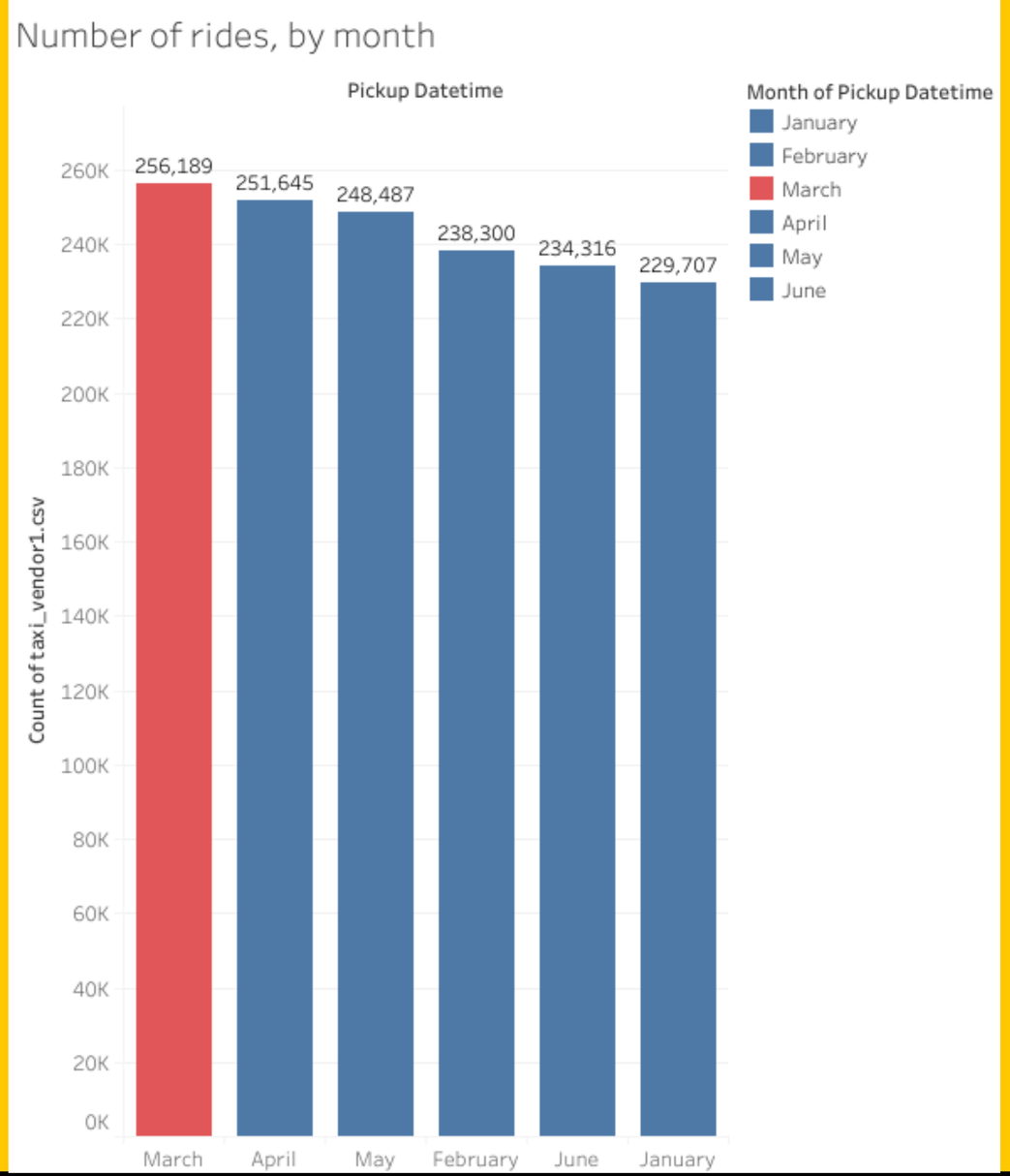
The Break Down

BASICS

- January 2016 - June 2016
- 1,458,644 trips !
- Vendor 1: 678,342 trips
- Vendor 2: 780,302 trips

POINTERS

- March - 256,189 trips
- Top 3 months - Spring
- Bottom 3 months - Winter



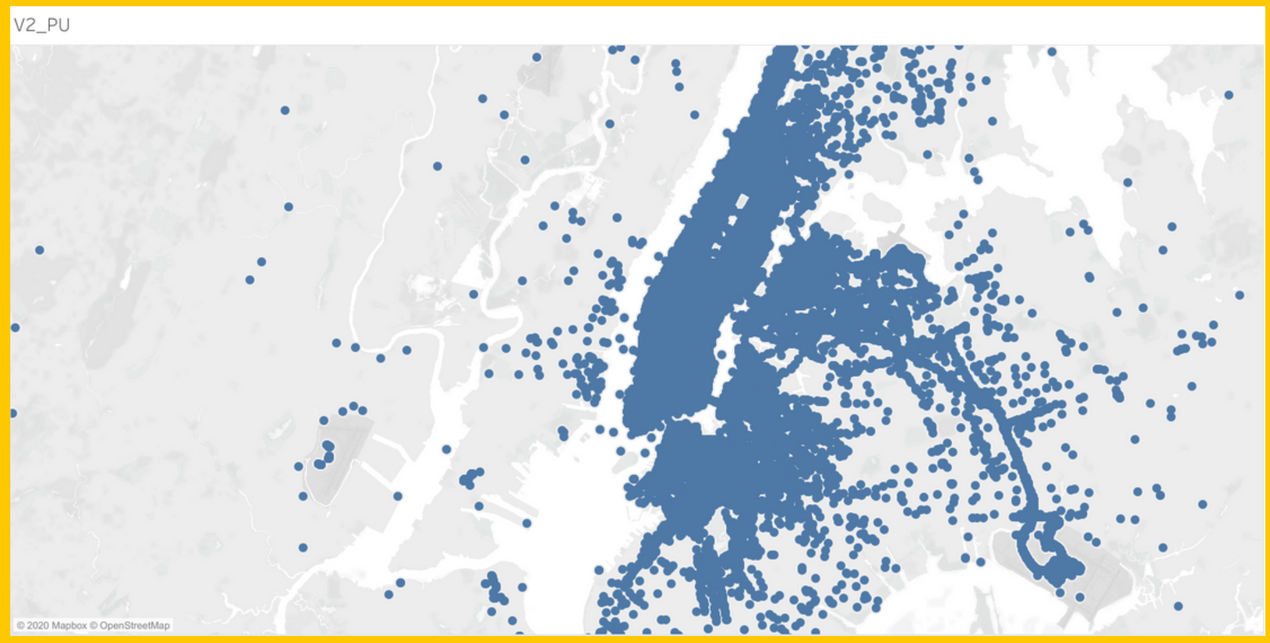
INSIGHT #1

FOCAL POINT

MANHATTAN IS A HIGH TRAFFIC AREA !

OTHER AREAS OF FOCUS

- NORTHERN BROOKLYN
- NORTHERN QUEENS
- AIRPORTS (JFK, LGA, EWR)



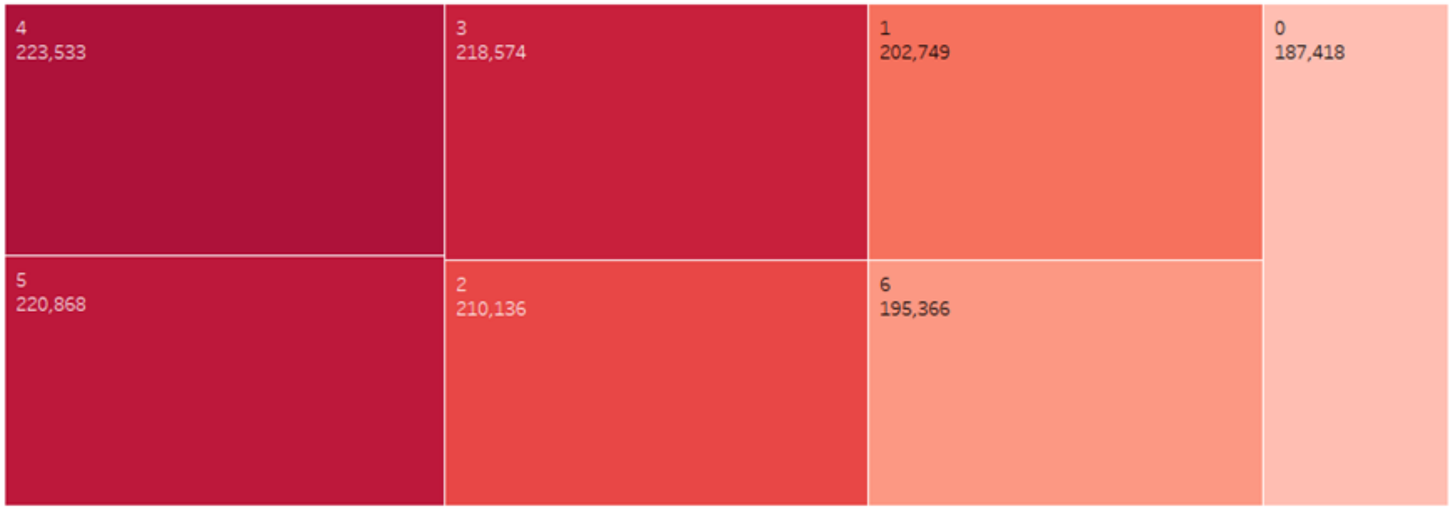
INSIGHT# 2

Friday (4) is the busiest day &
6pm (18:00) peak travel time.

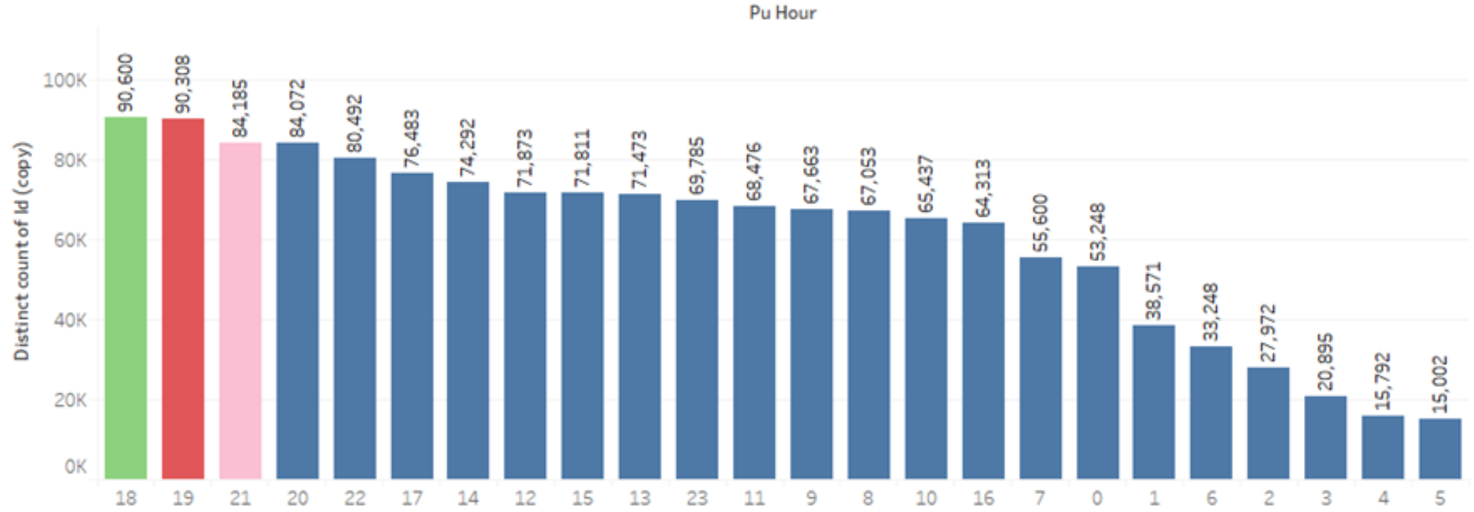
Points of Interest

Friday (4), Saturday
(5),& Thursday (3)
6 pm (18), 7pm(19) & 9pm(21)

Number of rides, by weekday

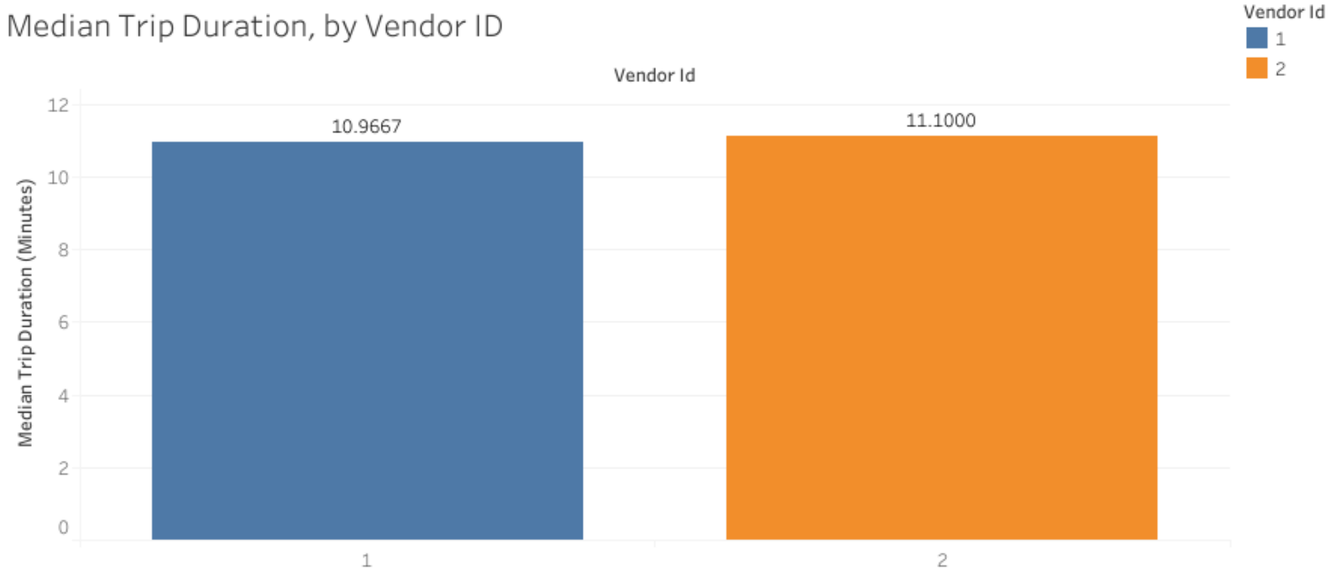


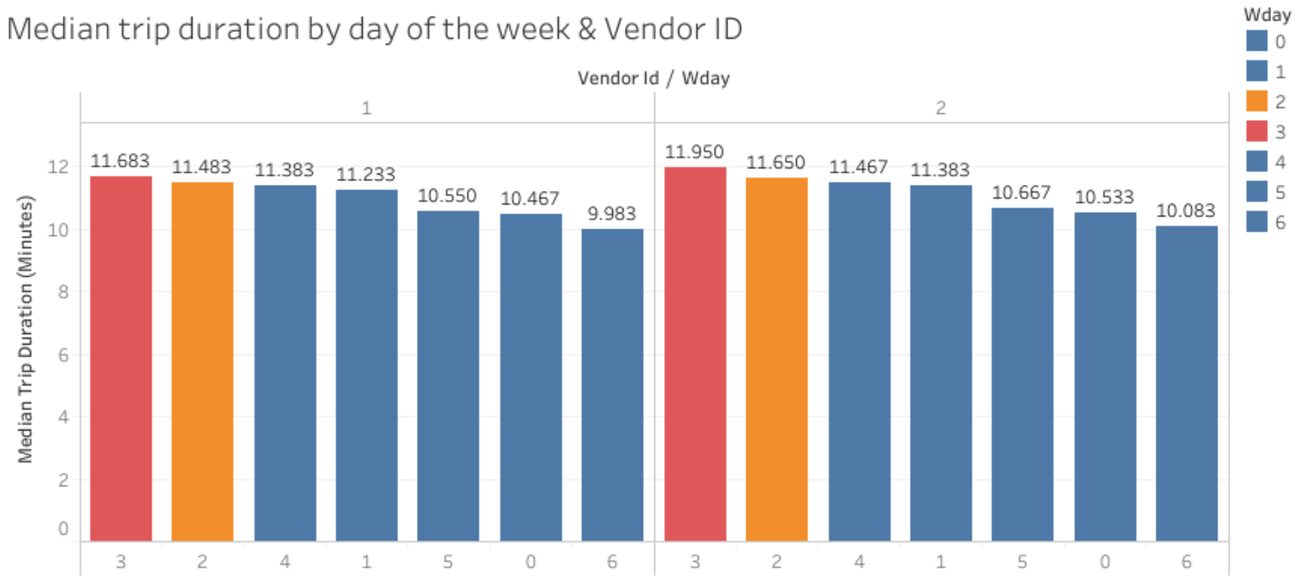
Number of rides, by pickup hour



INSIGHT# 3

Median Trip Duration for both vendors are roughly the same.





ARE WE THERE YET

PREDICTING

Trip Duration

HOW

1. Passenger Count
2. Pickup coordinates
3. Pickup hour
4. Day of the week
5. Month

RESULTS

WE **CAN NOT** PREDICT TRIP DURATION

RECOMMENDATIONS

- Different Model
- More Data

CONCLUSION

PRIMARY

- Different Model
- More Data

SECONDARY

- Manhattan
- Peak day & time
- Median Trip Duration

NEXT STEPS

- **Fare**
- **Distance**

APPENDIX

Out[9]:

OLS Regression Results

Dep. Variable:	trip_duration	**R-squared:**	0.000
Model:	OLS	**Adj. R-squared:**	0.000
Method:	Least Squares	**F-statistic:**	12.35
Date:	Mon, 28 Sep 2020	**Prob (F-statistic):**	1.55e-27
Time:	18:49:02	**Log-Likelihood:**	-1.1027e+07
No. Observations:	1093983	**AIC:**	2.205e+07
Df Residuals:	1093969	**BIC:**	2.205e+07
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	776.6557	24.461	31.751	0.000	728.714	824.597

Link to Dataset below

New York City Taxi Trip Duration

Share code and data to improve ride time predictions

k kaggle

EMAIL

danielwallace@ymail.com

LINKEDIN

https://www.linkedin.com/in/danielwallace1027/

GITHUB

https://github.com/GoldenTimbs

THANK YOU!