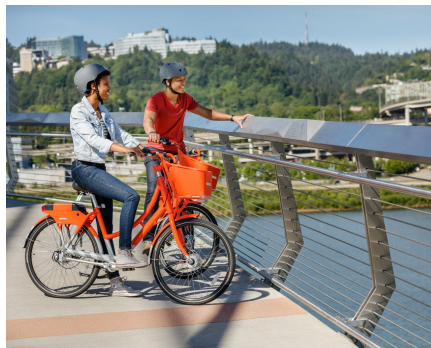
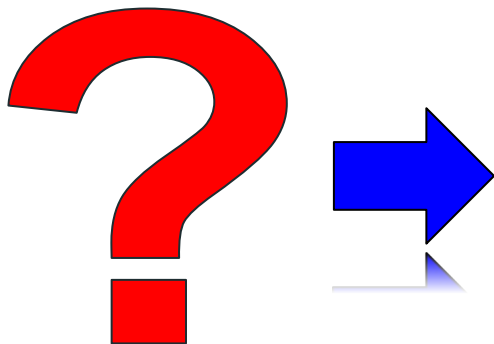


Predicting Ridership DC Capital Bikeshare

Joe McAllister and Mindy Zhou



Business Overview:



Number of total rentals
Number of casual users



Dataset:

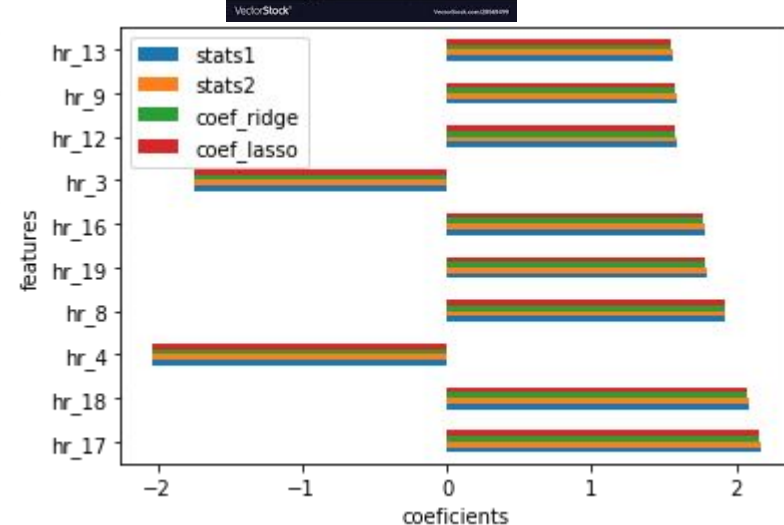
- UCI capital bike sharing dataset :
 - Number of rentals every hour for two years (2011 - 2013)
 - 17379 rows and 15 features

	season	yr	mnth	hr	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
dteday															
1-01-01	1	0	1	0	0	6	0	1	0.24	0.2879	0.81	0.0	3	13	16
1-01-01	1	0	1	1	0	6	0	1	0.22	0.2727	0.80	0.0	8	32	40
1-01-01	1	0	1	2	0	6	0	1	0.22	0.2727	0.80	0.0	5	27	32

Key Factors for Total Rentals

	stats1	stats2	ridge	lasso
0	weathersit_mist	holiday_not_holiday	weathersit_mist	weekday_Sat
1	weekday_Sat	weathersit_mist	weekday_Sat	workingday_working
2	workingday_working	weekday_Sat	workingday_working	weathersit_mist
3	weekday_Sun	workingday_working	weekday_Sun	weekday_Sun
4	season_summer	weekday_Sun	season_summer	season_summer

- ❑ Hours of the day is most important
 - ❑ High during
 - ❑ 8 - 9 am, 4 - 7 pm, ,12-1 pm
 - ❑ Low during:
 - ❑ 3 - 4 AM
- ❑ Least Important:
 - ❑ Misty day
 - ❑ Weekend
 - ❑ Workday or holiday
 - ❑ Summer time



R2 for test and train for all models:
~ 0.82

When are more non-members on the road?



Top 10 predictors (log regression)

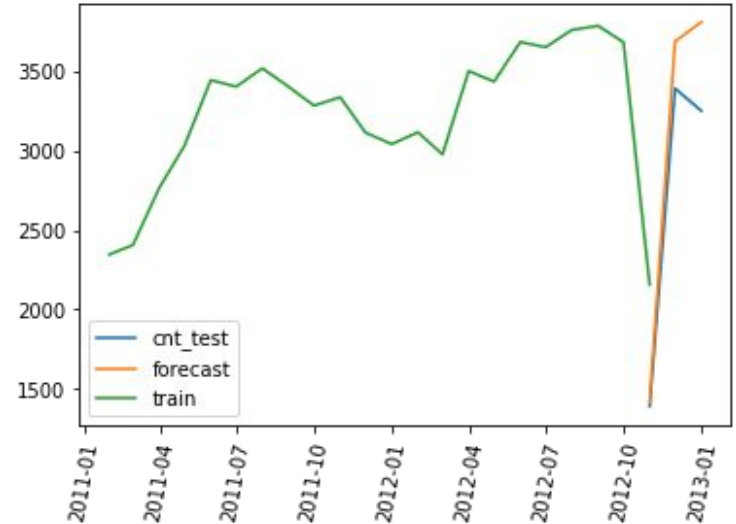
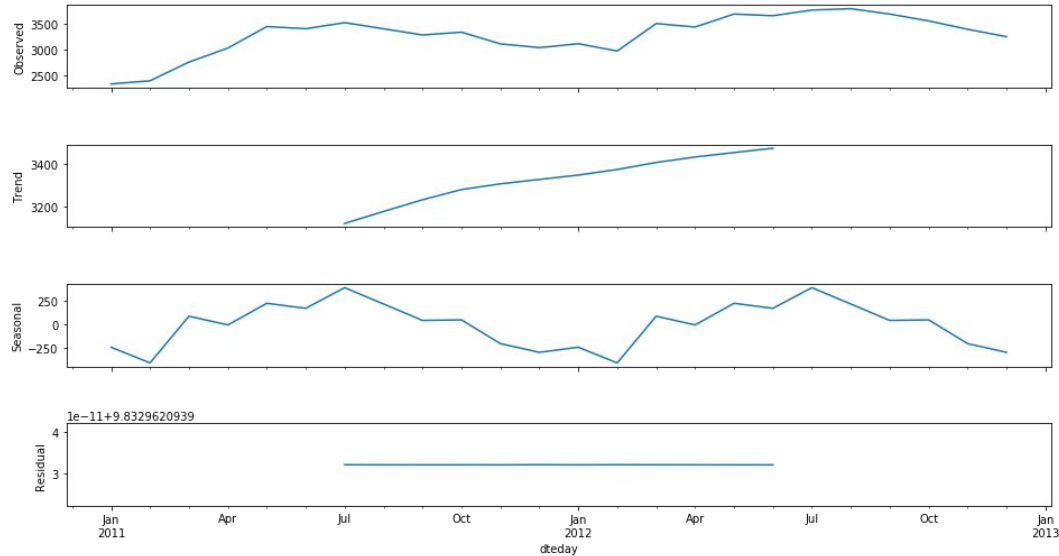
	predictor	log_coefs	transformed_coefs
19	hr_2	2.570454	13.071752
20	hr_3	2.560253	12.939085
21	hr_4	2.303639	10.010547
3	temp	1.753863	5.776877
0	holiday	1.552573	4.723610
18	hr_1	1.170523	3.223678
13	weekday_1	0.557421	1.746164
32	hr_15	0.426943	1.532565
31	hr_14	0.372294	1.451060
9	weathersit_1	0.362031	1.436243

Having mostly casual riders is a 'rare event'
~ 2.2% of the time

- ❑ Hour of day highest contributor
 - ❑ 2-4 AM
- ❑ Followed by
 - ❑ Temperature -> warmer higher likelihood
 - ❑ Holiday -> tourist influx?

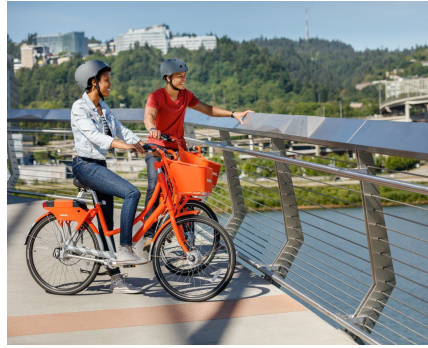
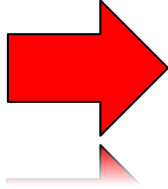
Note: AUC High (.82), but false positive classification is also high for non members

Trend, Seasonality, Forecast



- Trend (require more data):
 - Total bike rental increases with time
- Yearly seasonality:
 - Increase from January to July and decrease from July to December (match weather pattern)
- Forecast:
 - Not enough data to accurately forecast the downward projection of winter season

Conclusions:

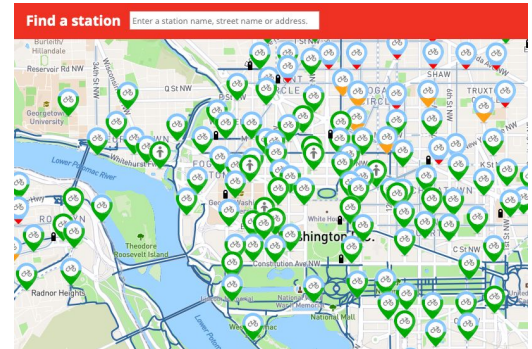


Number of total rentals
Number of casual users



Follow-ups:

- Gather more years of data
- Gather location information



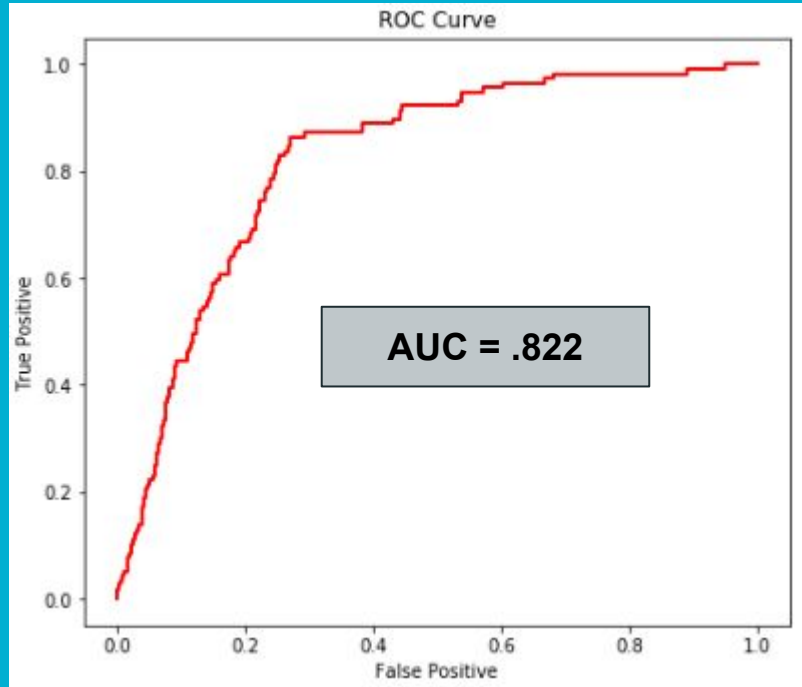
Thank You & Questions?

Appendix

References:

dataset: <https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset>

Logistic regression metrics



- ❑ Word of warning
 - ❑ High AUC can be misleading for rare event classification

Confusion Matrix

Actual 0	4441	656
Actual 1	54	63
	Predict 0	Predict 1