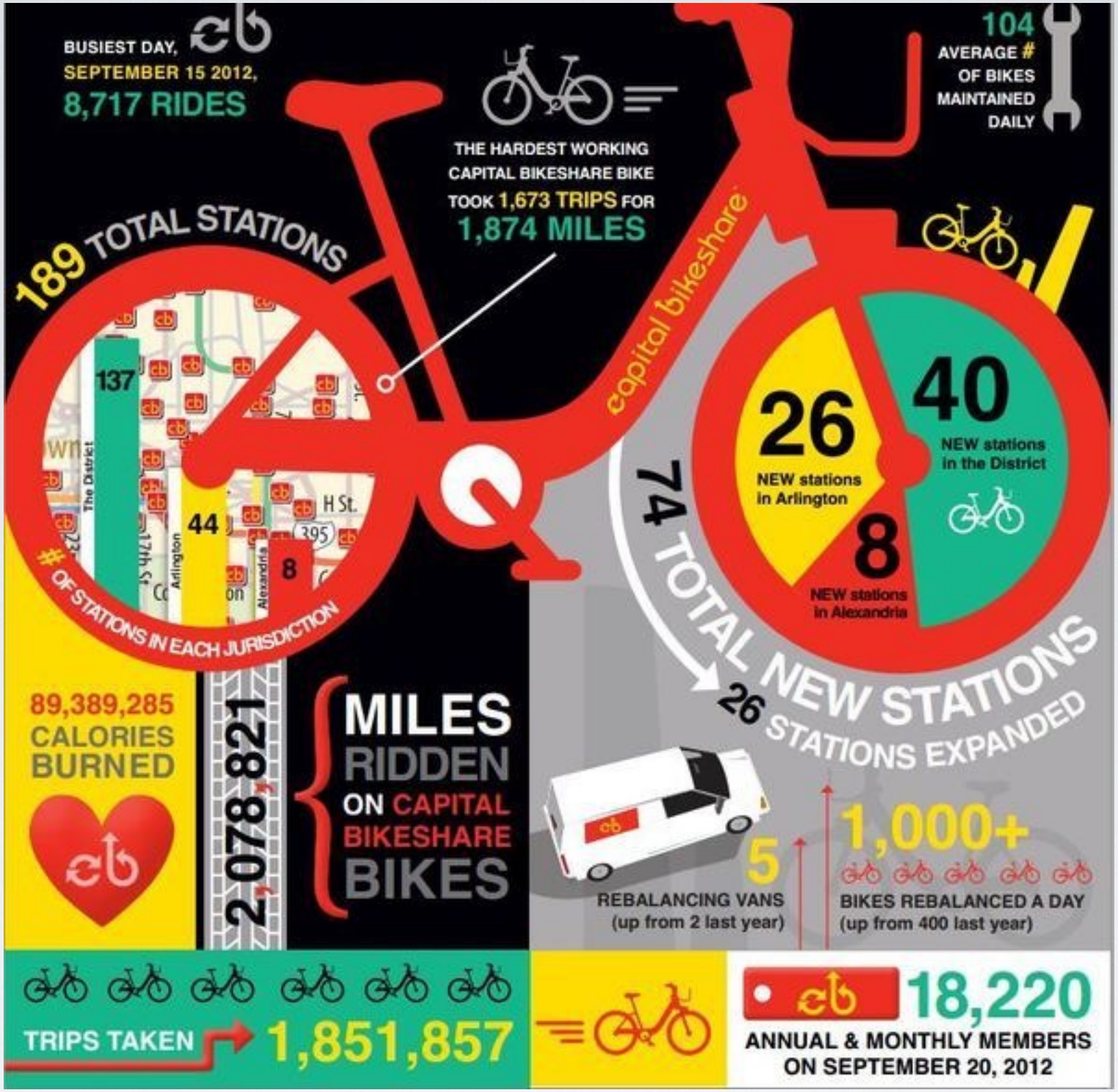


DC BIKE SHARE DEMAND

A Kaggle Competition

THE CHALLENGE

1. Forecast hourly Capitol Bikeshare demand given information about weather and features of time.

2. Fit the model on "training data" (first 19 days of the month) and apply it to "test" data (remaining days).

   - This is different than what we covered in class–the test data is actually what we called "new data" (it does not contain data on demand)

3. Submit data to Kaggle for a score:

   - Kaggle compares fitted data on the "test" dataset to actual demand, and calculates a RMSLE (Root Mean Squared Log Error)

THE DATA AS DOWNLOADED:

(ALREADY VERY CLEAN)

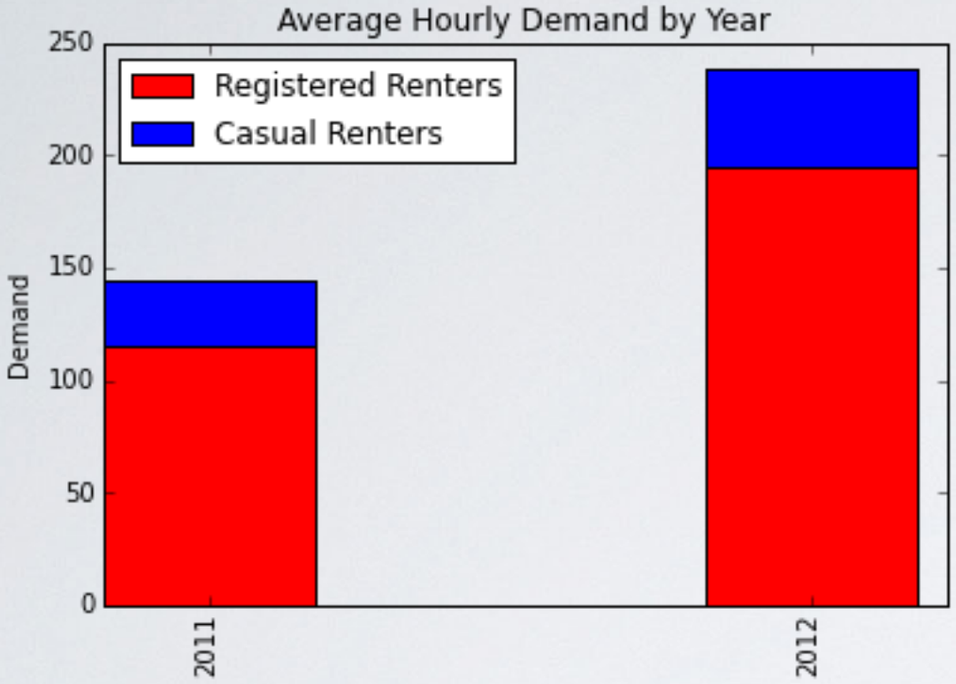
Time	**Weather**	**Demand**
Y/M/D/H	Season	Registered
Holiday?	Conditions	Casual
Workday?	Temperature (2)	
	Windspeed	
	Humidity	

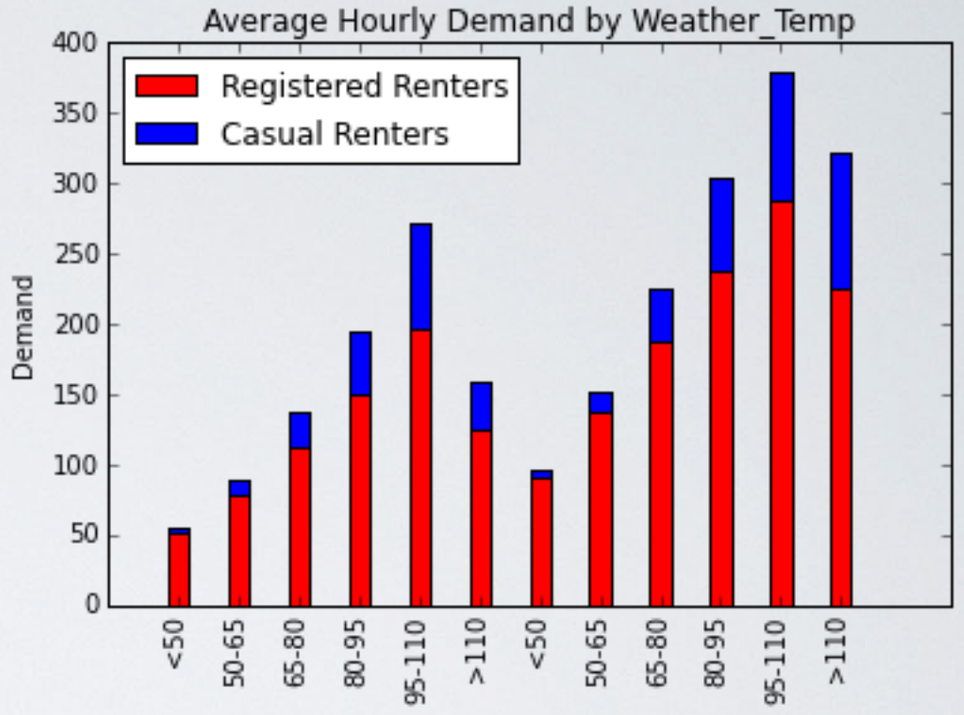
THE DATA TRANSFORMED:

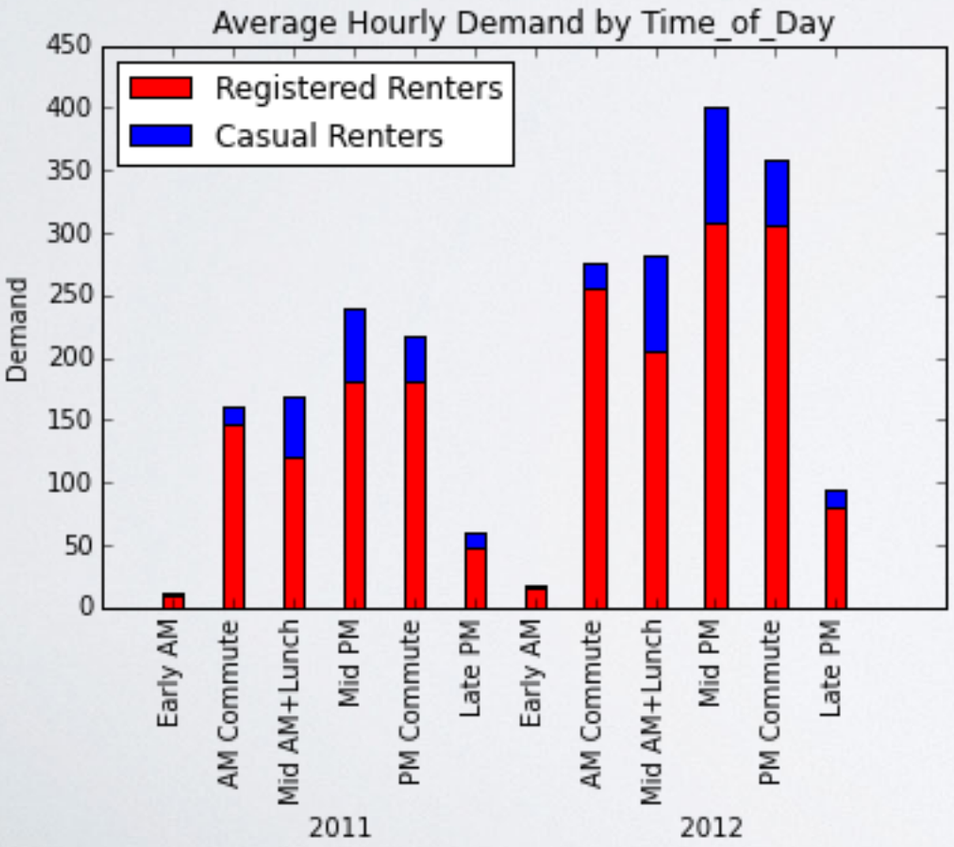
(A FEW MORE FEATURES)

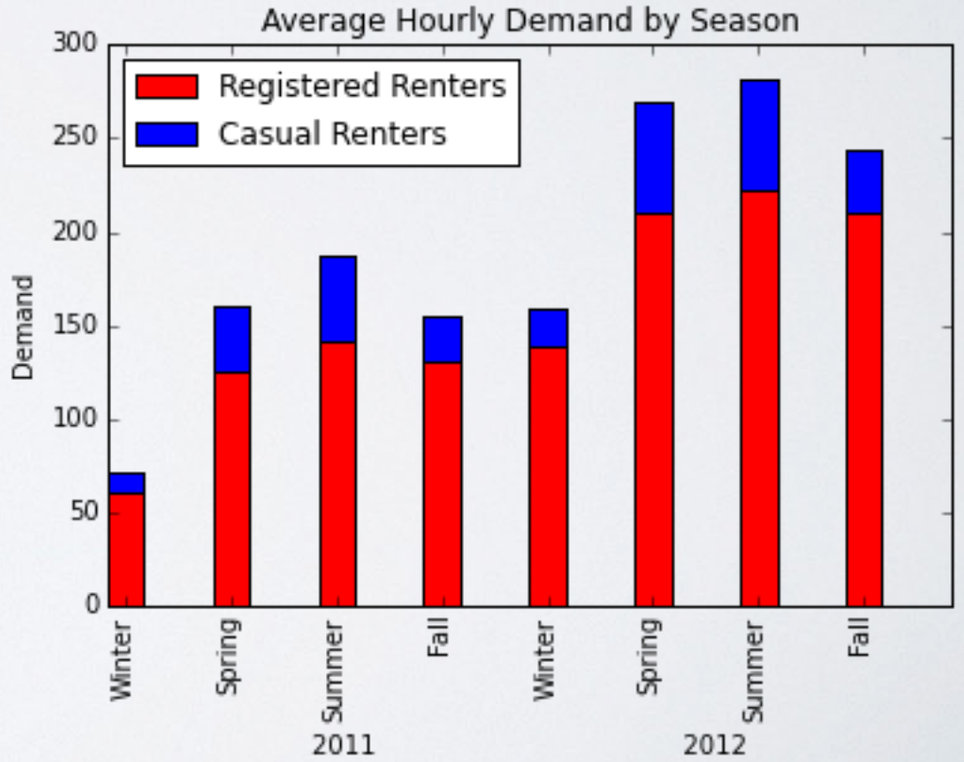
Time	**Weather**	**Demand**
Y,M,D,H	Season	Registered
Holiday?	Conditions	Casual
Workday?	Temperature (2)	
	Windspeed	
	Humidity	

THE DATA IN GRAPHS:









GOING FORWARD

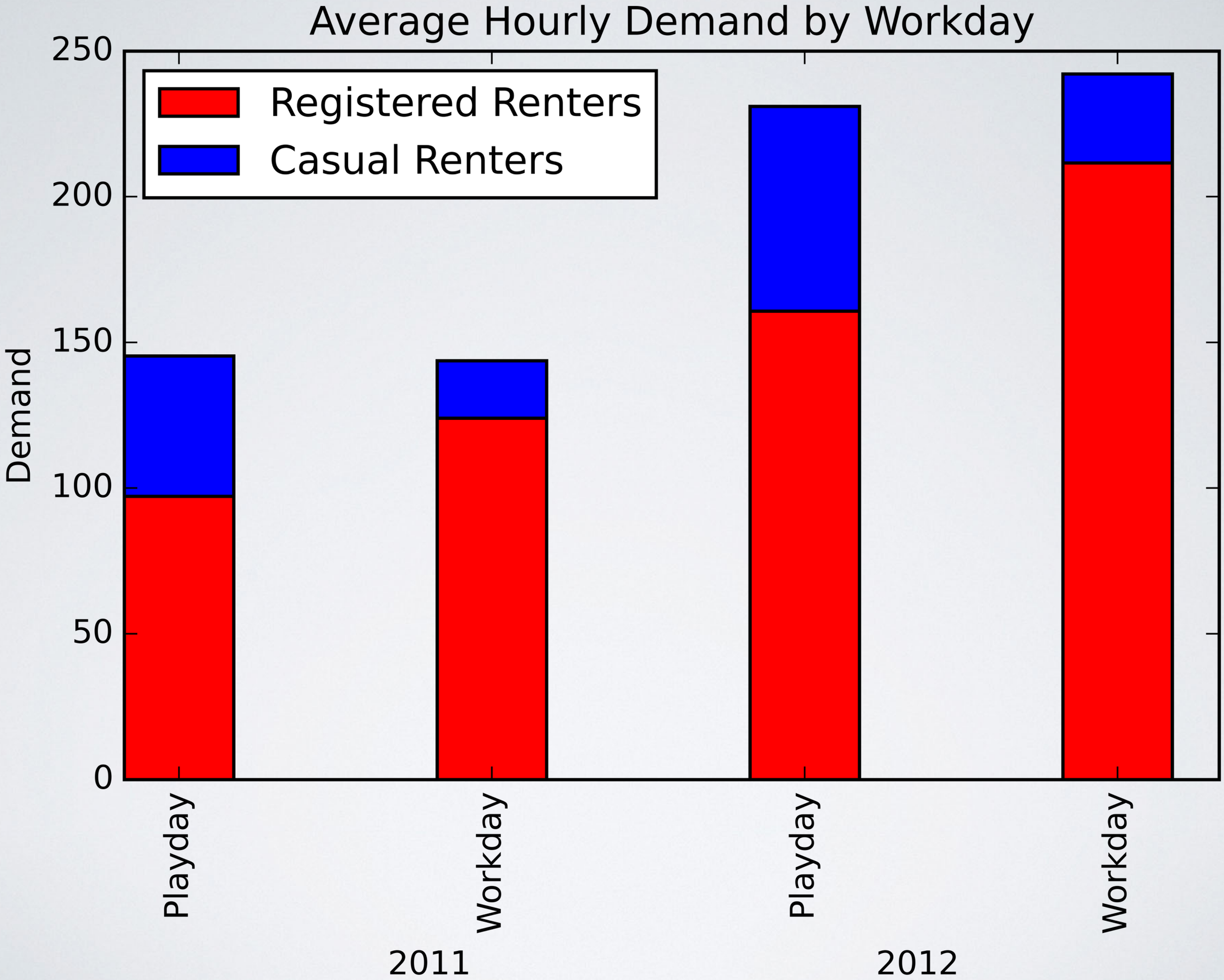
- I will use a **regression model** to predict demand in the "test" dataset

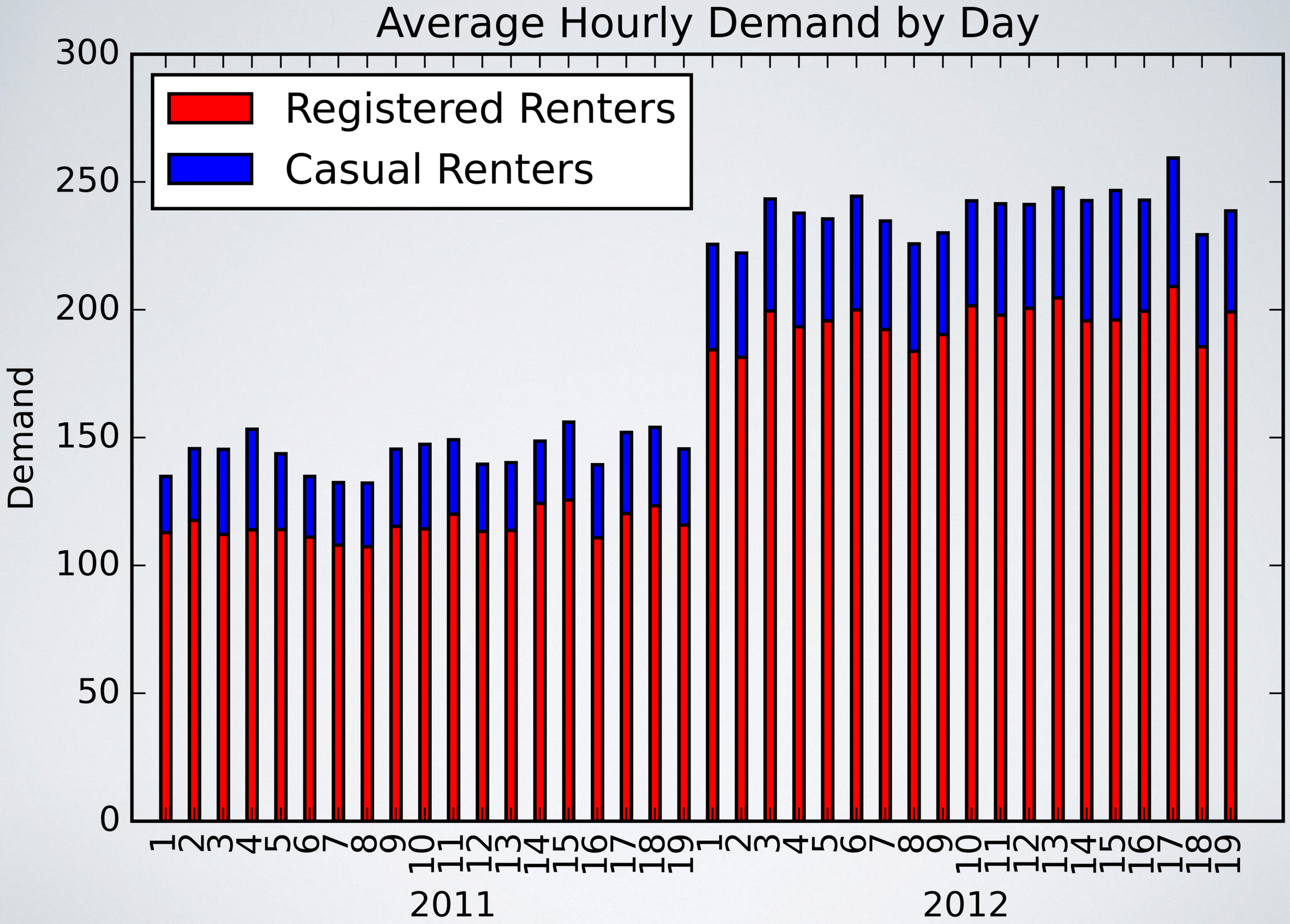
- People are ranking in the 15% of the competition by using just using the data available (randomForest).

    - **But,** the rules do not state other data can cannot be used to make predictions.

        - WMATA has an API on metro rail/bus disruptions.

RESERVE SLIDES


Average Hourly Demand by Workday
Registered Renters
Casual Renters
Demand
250
200
150
100
50
0
Playday
Workday
Playday
Workday
2011
2012


Average Hourly Demand by Day
Registered Renters
Casual Renters
Demand
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 3 14 15 16 17 18 19 1 2 3 4 5 6 7 8 9 10 11 12 3 14 15 16 17 18 19
2011
2012

Average Hourly Demand by Hour

Registered Renters

Casual Renters

Demand

600

500

400

300

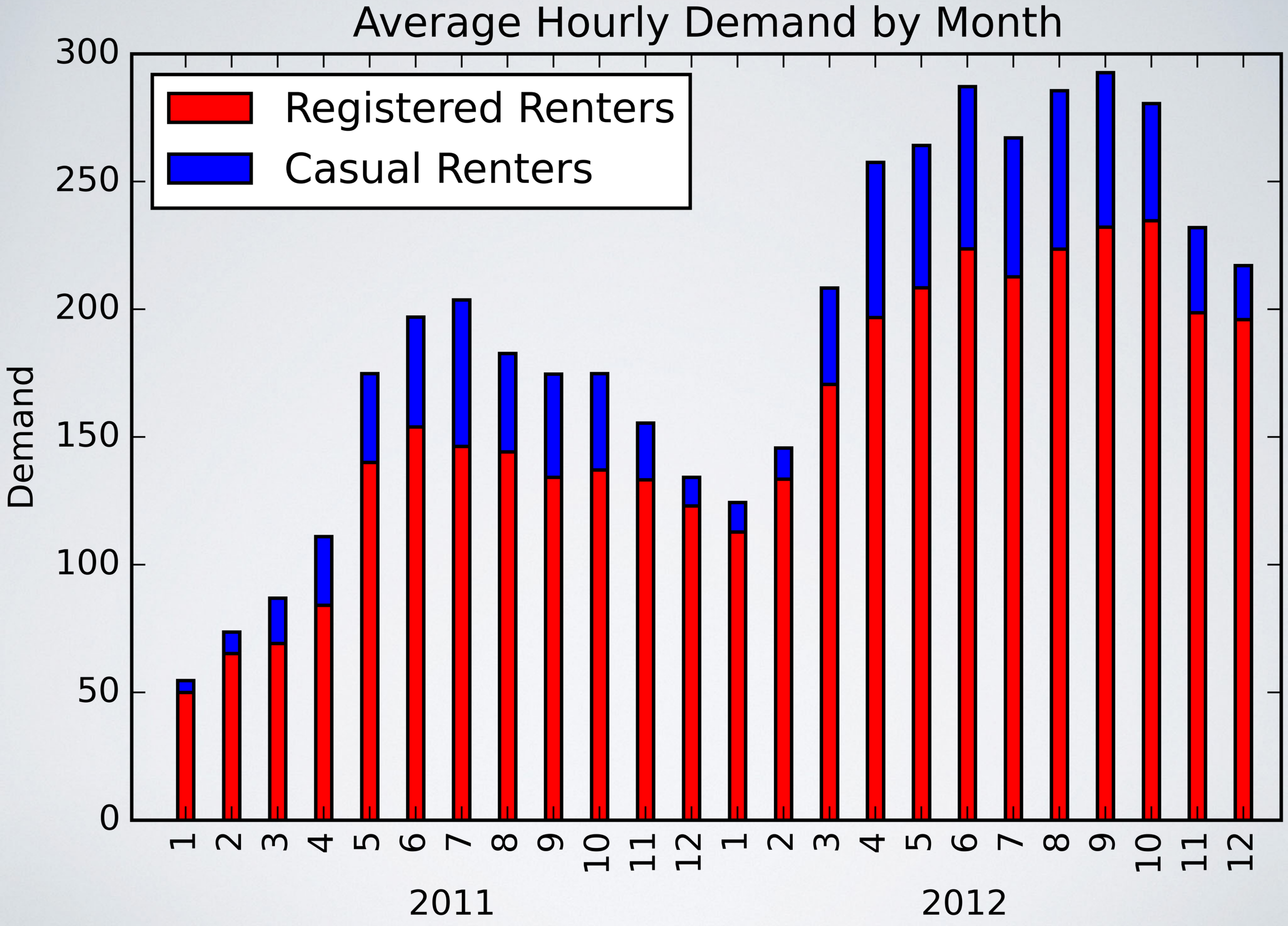
200

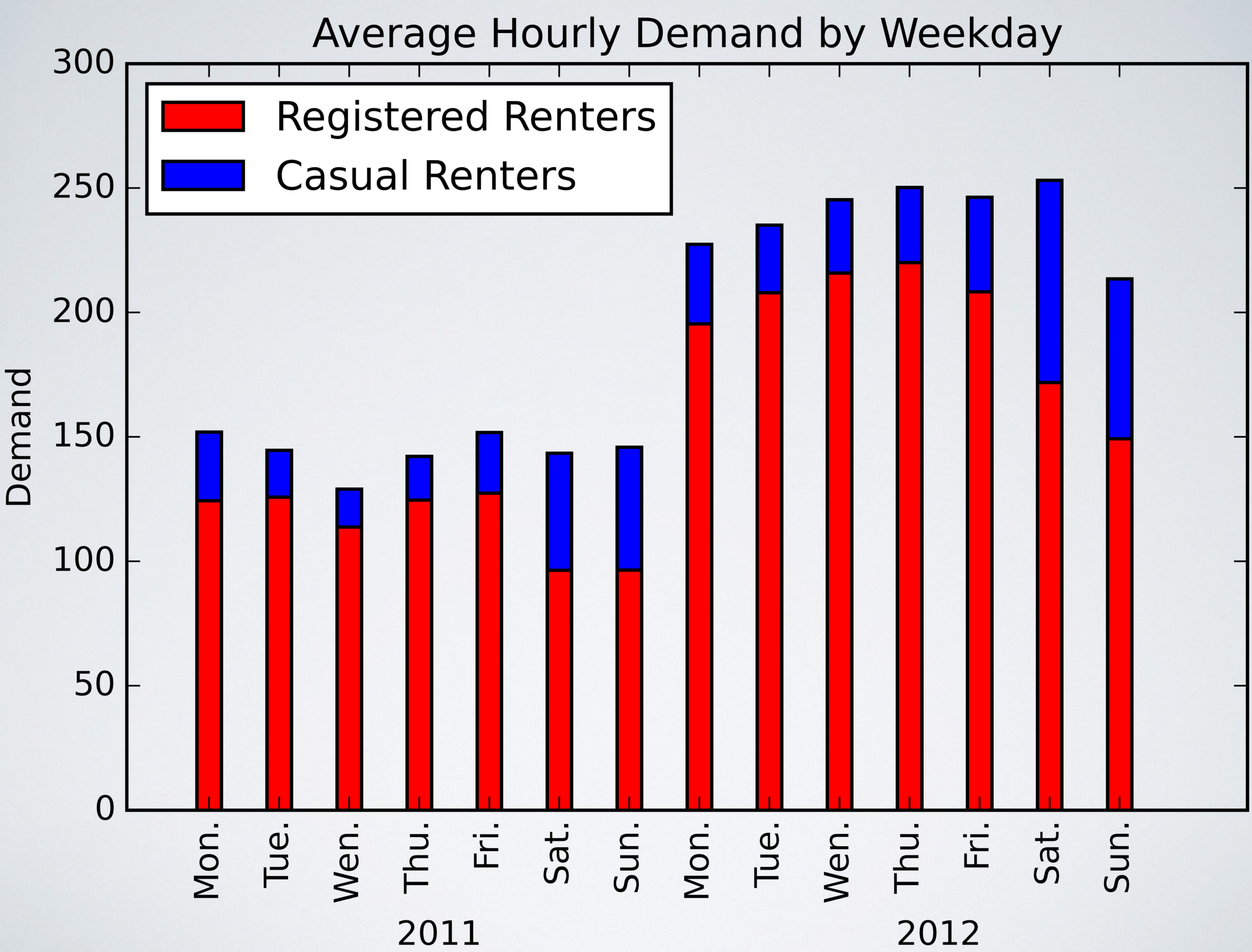
100

0

2011

2012


Average Hourly Demand by Month
Registered Renters
Casual Renters
Demand
300
250
200
150
100
50
0
1
2
3
4
5
6
7
8
9
10
11
12
1
2
3
4
5
6
7
8
9
10
11
12
2011
2012


Average Hourly Demand by Weekday
Registered Renters
Casual Renters
Demand
300
250
200
150
100
50
0
Mon.
Tue.
Wen.
Thu.
Fri.
Sat.
Sun.
Mon.
Tue.
Wen.
Thu.
Fri.
Sat.
Sun.
2011
2012