Bike Sharing Demand Analysis

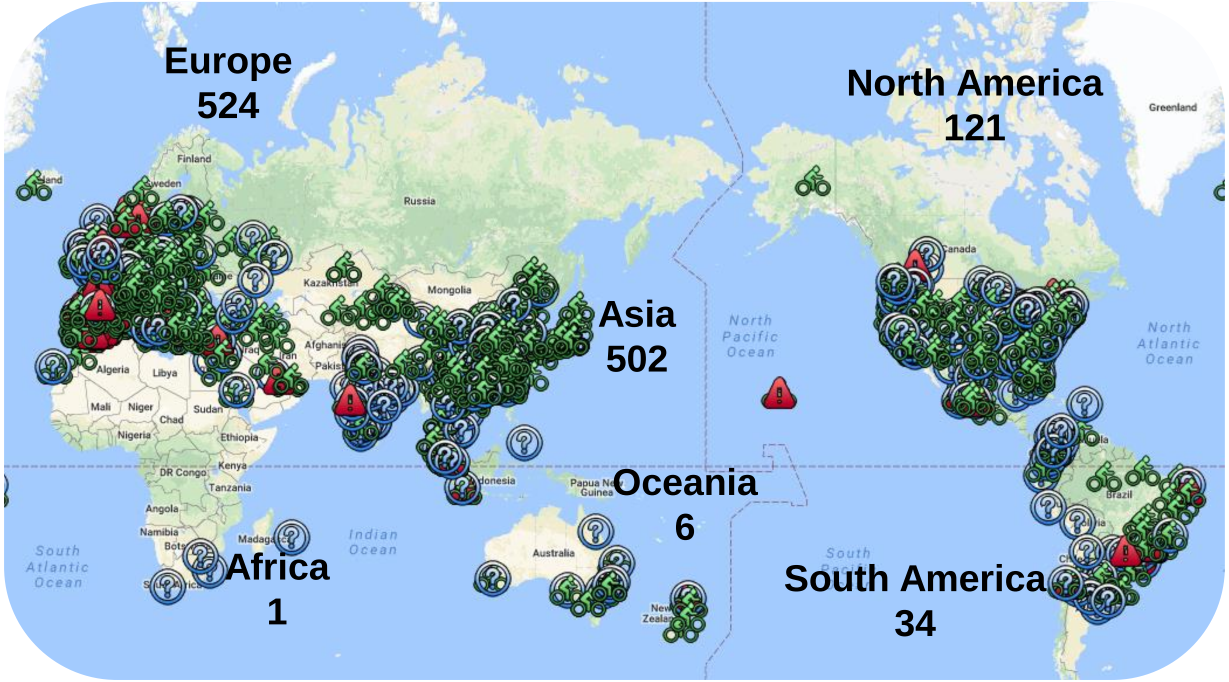
Min Fang

Outline

- ❑ Introduction
- ❑ Model building
- ❑ Results and conclusion
- ❑ Future work and Applications

Bike Sharing System

- New generation of traditional bike rental (4th generation)
- Provide self-service point-to-point short trips
- Benefits:
  - reduce traffic congestion
  - mitigate pollution
  - health benefits
  - greater environmental awareness
- Global expansion



adapted from www.bikesharingmap.com

Data Description

- Our data set was obtained by the UC Irvine Machine Learning Repository
  (http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset)

- hour.csv file was selected
  hourly count of rental bikes

- **17,379** observations

- **13** predictior variables
  - **9 Indicator predictors:** record index, date, season, year, month, hour, holiday, weekday, working day, weather condition
  - **4 Quantitative predictors:** temp, feeling temp, wind speed and humidity

- **3** response variables
  - counts from casual users, registered users and total counts

Data and Variable Treatment

- Data treatment
  - Added a new dummy variable “type”: casual is 1, registered is 0

- Variable treatment
  - Indicator variables (season, mnth, weekday, holiday, workingday, weathersit, yr, and hr) were converted into factors
  - The variable hour was divided into five levels

(hours 1 [0-5], hours 2 [6-9], hours 3 [10-15], hours 4 [16-20], hours 5 [21-23])

Motivations

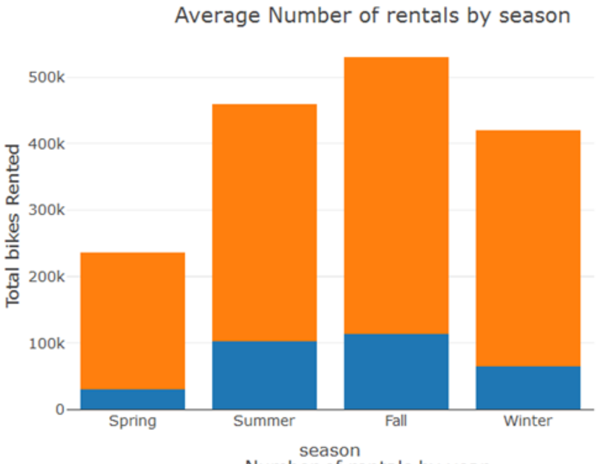
- Problem: There exist different bike sharing demand for casual and registered users
    - different needs and difficulties
- Our goal: To describe the differences on the bike sharing demand between casual and registered users
    - To be more specific, we want to build a regression model to describe bike sharing demand for a specific group in a given scenario
- Benefits: Government or bike provider can provide better service and deal with different needs of subgroups of users

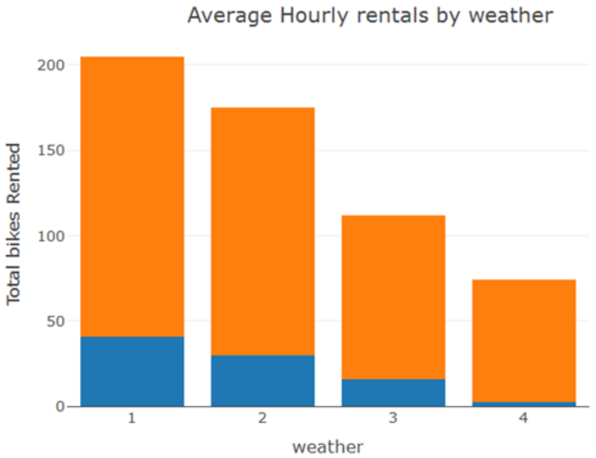
Research Questions

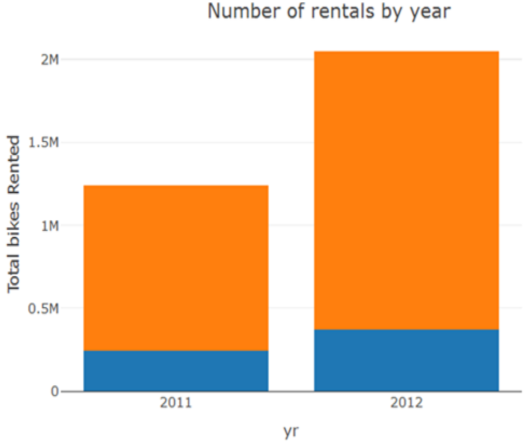
- ❑ Which predictor variables significantly influenced bike rental demand?
- ❑ Did the predictors affect bike rental demand differently in casual and registered user groups?
- ❑ Can we fit a model to describe such differences?

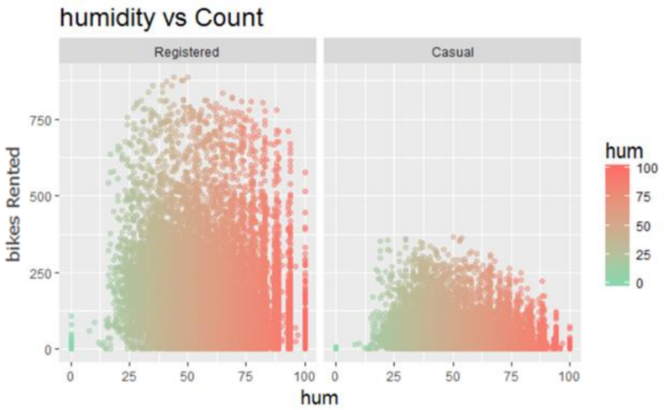
Significant Predictors (season / yr / weather / hum)

Registered
Casual

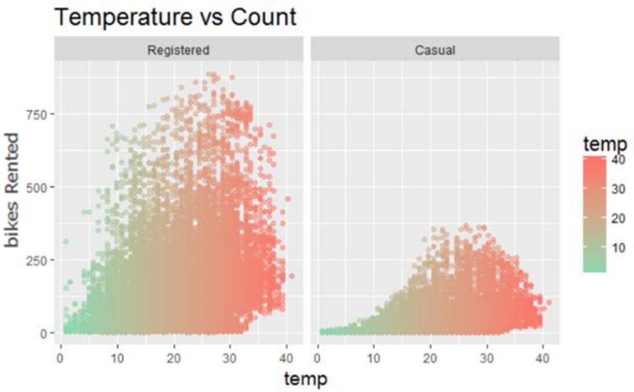


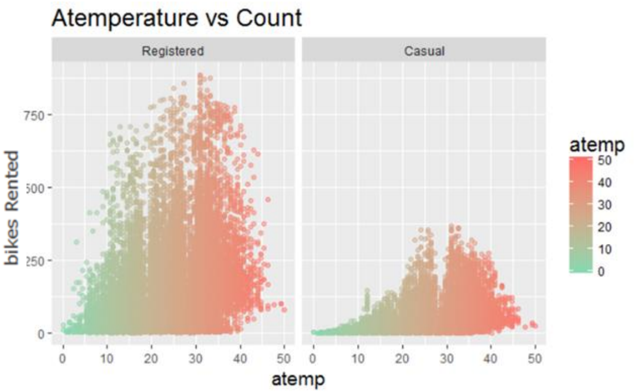


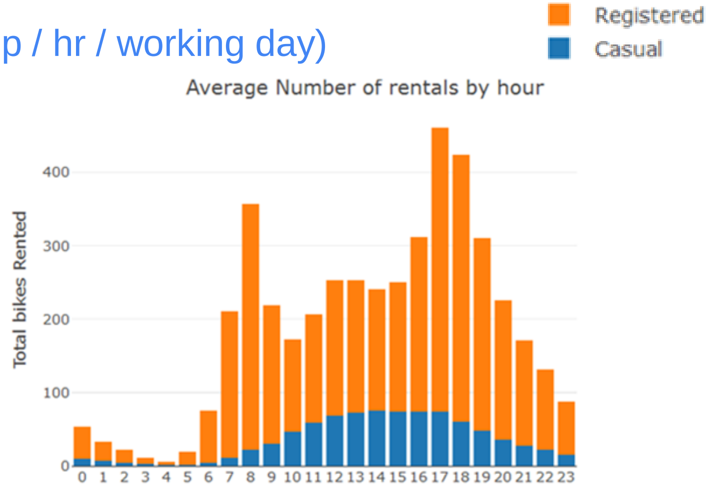




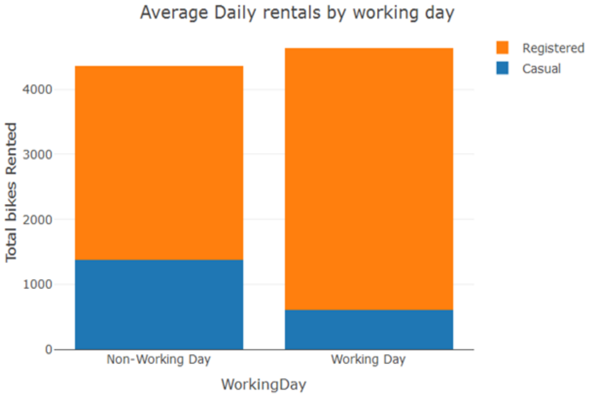
Significant Predictors (temp / atemp / hr / working day)







Type Matters (type ~ hr ~ workingday)

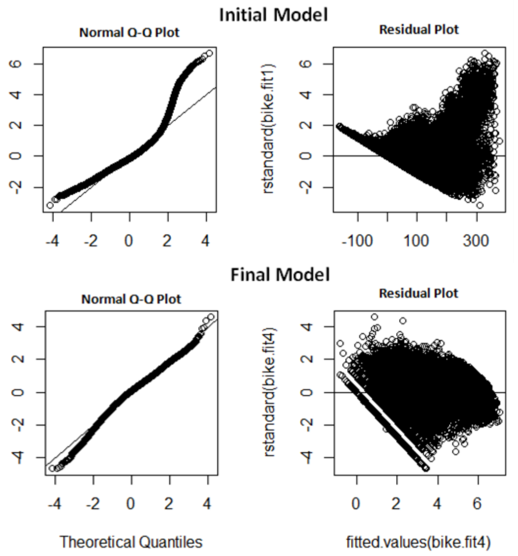


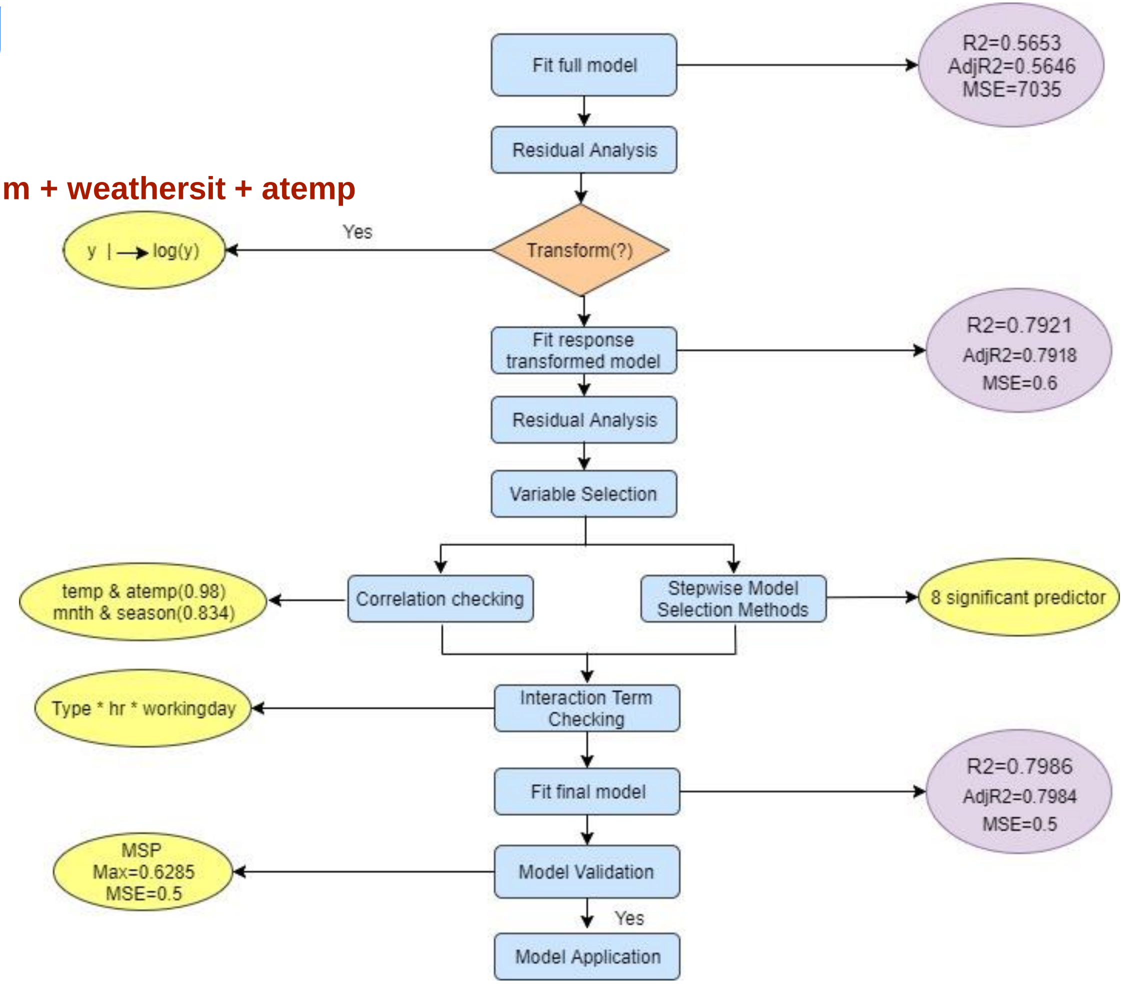
Model Building

Final model :

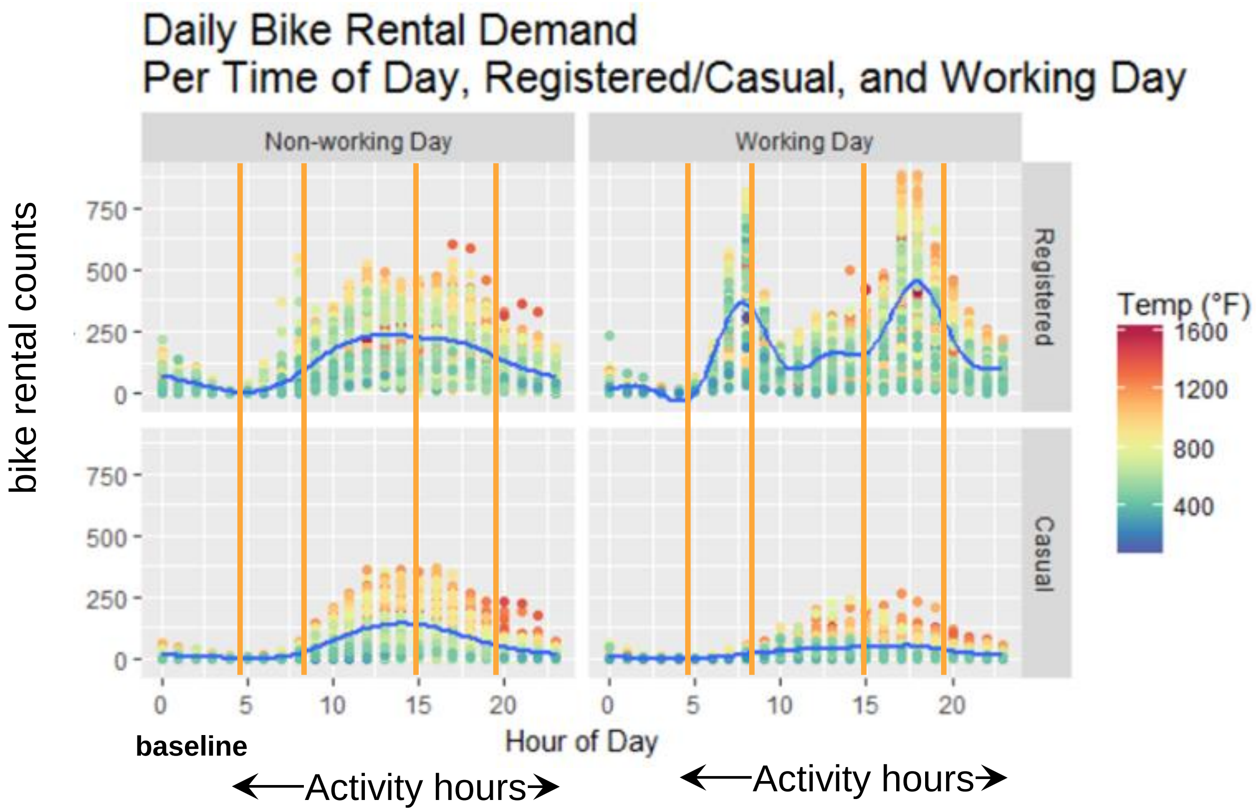
log(user) ~

season + yr + type*workingday*hr + hum + weathersit + atemp





Can the model describe the distinct patterns observed?



Translate model estimates into the rental counts ratio

The ratio is the bike rental counts during four activity hours when compared to the baseline hours [0-5].

	Ratio (activity hr/baseline hr) exp(estimated parameters)			
	Working day		Non-Working day	
Activity hr	casual	registered	casual	registered
Morning peak hr (hr2)	4.415	24.095	1.941	1.941
Middle of the day(hr3)	8.776	10.591	11.905	7.330
Evening peak hr (hr4)	9.143	24.631	7.404	3.367
Night time (hr5)	4.637	10.166	2.869	3.367

Groups of interest (casual/registered/workingday)

↓

Make contrasts

↓

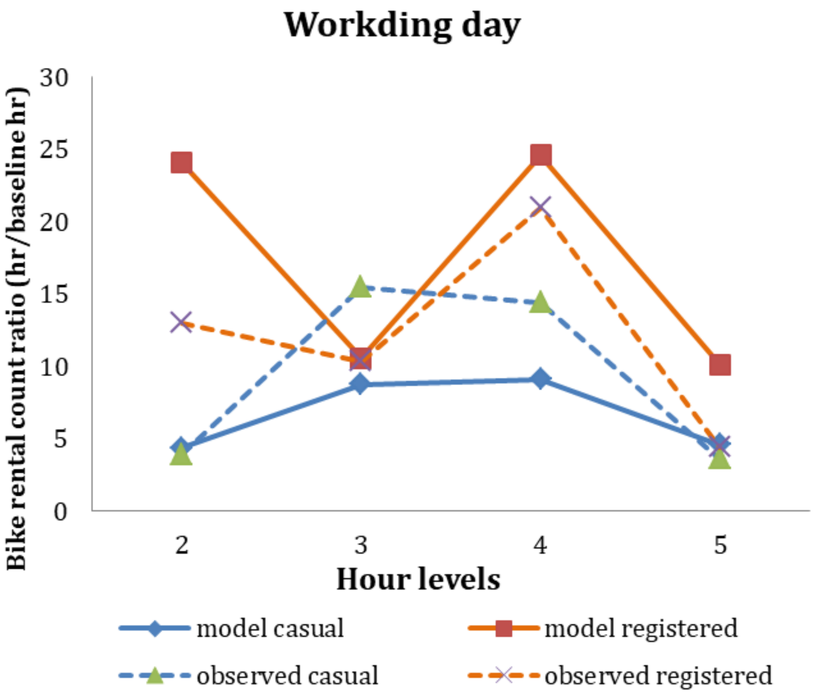
Find corresponding estimated parameters

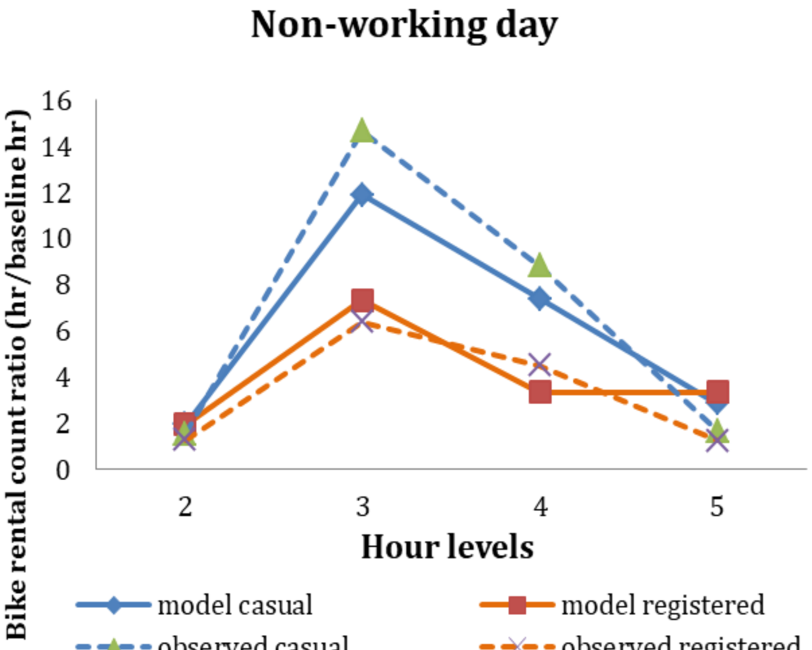
↓ **Δlog(activity hr/baseline hr)**

Ratio = exp (estimated parameters)

Ratio: hr2 / hr1
hr3 / hr1
hr3 / hr1
hr4 / hr1

Bike rental patterns described by the model agreed with observed patterns





Conclusion

- We fitted a linear regression model to describe the different effect of various predictors on the bike rental counts from casual and registered users.
- The addition of a three-way interaction of important predictors (hr*type*workingday) greatly improved the model fitting.
- The model successfully described the patterns of bike rental counts against four hour levels from casual and registered users. And the patterns differed significantly on working and non-working days.

Future work & Applications

- ❑ Althernative model fitting

    Generalized linear model with Poisson distribution

    Nonparametric model using random forest method

- ❑ Model applications

    Predict bike sharing demand in a given scenario

    Optimaize bike allocation

Improve on-demand user experience (e.g. helmet rental, payment instruction and etc.)

Translate model estimates into the rental counts ratio

	Estimated parameters **∆log(hr2-5/hr1)**				**Ratio (hr2-5/hr1)** **exp(estimated parameters)**			
	Working day		Non-Working day		Working day		Non-Working day	
hr comparison	casual	registered	casual	registered	casual	registered	casual	registered
Morning peak hr (hr2) vs. midnight (hr1)	1.485	3.182	0.663	0.663	4.415	24.095	1.941	1.941
Middle of the day(hr3) vs. midnight (hr1)	2.172	2.36	2.477	1.992	8.776	10.591	11.905	7.330
Evening peak hr (hr4) vs. midnight (hr1)	2.213	3.204	2.002	1.214	9.143	24.631	7.404	3.367
Night time (hr5) vs. midnight (hr1)	1.534	2.319	1.054	1.214	4.637	10.166	2.869	3.367