Summary

We use the following scerntific calculation tools to finish my project

> numpy:
>
> Sklearn
>
> pandas

I Learned the use of git command from the project.

Code flow

1.Read the data set and build features

```
X_df= data.iloc[:,6:13]
X_df
```

	weekday	workingday	weathersit	temp	atemp	hum	windspeed
0	6	0	2	0.344167	0.363625	0.805833	0.160446
1	0	0	2	0.363478	0.353739	0.696087	0.248539
2	1	1	1	0.196364	0.189405	0.437273	0.248309
3	2	1	1	0.200000	0.212122	0.590435	0.160296
4	3	1	1	0.226957	0.229270	0.436957	0.186900
...	...	...	...	...	...	...	...
726	4	1	2	0.254167	0.226642	0.652917	0.350133
727	5	1	2	0.253333	0.255046	0.590000	0.155471
728	6	0	2	0.253333	0.242400	0.752917	0.124383
729	0	0	1	0.255833	0.231700	0.483333	0.350754
730	1	1	2	0.215833	0.223487	0.577500	0.154846

2.Build our LinearRegression and Random Forest model

```
regr = linear_model.LinearRegression()
regr.fit(train_X,train_y)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Random Forest

```
from sklearn.ensemble import RandomForestRegressor
```

```
regr_rf_day = RandomForestRegressor(max_depth=2, random_state=0)
regr_rf_day.fit(train_X,train_y)
regr_rf_day.score(test_X,test_y)
```

```
0.5381973663721356
```

```
regr_rf_hour = RandomForestRegressor(max_depth=2, random_state=0)
regr_rf_hour.fit(train_X_hour,train_y_hour)
regr_rf_hour.score(test_X_hour,test_y_hour)
```

```
0.21791703834431253
```

3.Result analysis

We use the coefficient of determination R^2 of the prediction to evaluate model

	Linear Regression	**Random Forest**
daily.csv	0.46	0.54
hourly.csv	0.26	0.22

4.Upload to github

https://github.com/hamy95/Bikeshare