

```
In [111]: import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.dummy import DummyRegressor
from sklearn.model_selection import train_test_split
```

```
In [112]: dodger = pd.read_csv('dodgers.csv')
dodger.head(10)
```

Out[112]:

	month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks	bobblehead
0	APR	10	56000	Tuesday	Pirates	67	Clear	Day	NO	NO	NO	NO
1	APR	11	29729	Wednesday	Pirates	58	Cloudy	Night	NO	NO	NO	NO
2	APR	12	28328	Thursday	Pirates	57	Cloudy	Night	NO	NO	NO	NO
3	APR	13	31601	Friday	Padres	54	Cloudy	Night	NO	NO	YES	NO
4	APR	14	46549	Saturday	Padres	57	Cloudy	Night	NO	NO	NO	NO
5	APR	15	38359	Sunday	Padres	65	Clear	Day	NO	NO	NO	NO
6	APR	23	26376	Monday	Braves	60	Cloudy	Night	NO	NO	NO	NO
7	APR	24	44014	Tuesday	Braves	63	Cloudy	Night	NO	NO	NO	NO
8	APR	25	26345	Wednesday	Braves	64	Cloudy	Night	NO	NO	NO	NO
9	APR	27	44807	Friday	Nationals	66	Clear	Night	NO	NO	YES	NO

```
In [113]: cleanup_nums = {"bobblehead": {"NO": 0, "YES": 1}}
dodger.replace(cleanup_nums, inplace=True)
```

Baseline Regression Model

```
In [114]: bobblehead, attend = dodger.attend, dodger.bobblehead

bobblehead_train, bobblehead_test, attend_train, attend_test = train_test_split(
    bobblehead, attend, random_state=0)
```

```
In [115]: dummy = DummyRegressor(strategy='mean')
dummy.fit(bobblehead_train.values.reshape(-1,1), attend_train.values.reshape(-1,1))
```

```
Out[115]: DummyRegressor(constant=None, quantile=None, strategy='mean')
```

```
In [116]: # Get R-squared score
dummy.score(bobblehead_test, attend_test)
```

```
Out[116]: -0.03480263157894736
```

To compare, we train our model and evaluate the performance score:

```
In [117]: ols = LinearRegression()
ols.fit(bobblehead_train.values.reshape(-1,1), attend_train.values.reshape(-1,1))
```

```
Out[117]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
In [118]: ## Get R-squared score
ols.score(bobblehead_test.values.reshape(-1,1), attend_test.values.reshape(-1,1))
```

```
Out[118]: 0.25345451634657756
```

```
In [121]: dodger.attend.min()
```

```
Out[121]: 24312
```

```
In [123]: dodger.query('attend == 24312')
```

```
Out[123]:
```

	month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks	bobblehead
18	MAY	14	24312	Monday	Snakes	67	Clear	Night	NO	NO	NO	0

```
In [ ]:
```

