In [150]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Loading and Pre-Processing Data

In [151]:
```
clubs = ["rudas", "oxfordsc_sd", "wetrepublic", "templeSF", "grandBoston", "grandSF", "prsymCH", "
data = pd.DataFrame()
for club in clubs:
    df = pd.read_csv(str("data_"+ club + "_likes.csv"))
    data = data.append(df)
```

Analysis of Gender

In [152]:
```
# % with gender successfully identified
genders = data['GENDER']
1 - genders.isnull().sum() / len(genders)
```

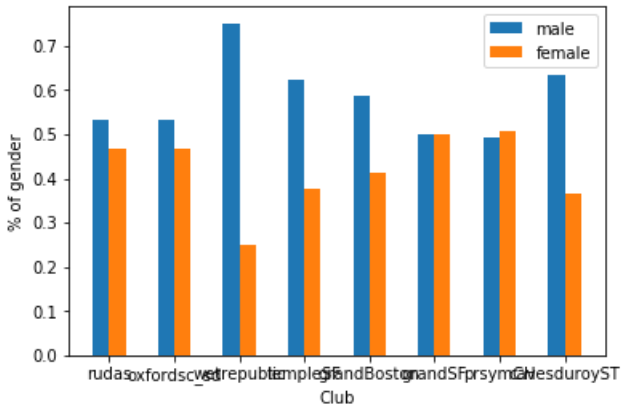
Out[152]: 0.6218214607754734

In [153]:
```
# % male and female likes respectively
genders = genders[genders.notnull()]
sum(genders == "male")/len(genders), sum(genders == "female")/len(genders)
```

Out[153]: (0.6387346702021875, 0.3612653297978124)

In [154]:
```
# gender breakdown by club
males = []
females = []
for club in clubs:
    df = pd.read_csv(str("data_"+ club + "_likes.csv"))
    males += [sum(df['GENDER'] == 'male')/sum(df['GENDER'].notnull())]
    females += [sum(df['GENDER'] == 'female')/sum(df['GENDER'].notnull())]
plt.bar(pd.Series(range(len(clubs)))-0.125, height = males, width = 0.25)
plt.bar(pd.Series(range(len(clubs)))+0.125, height = females, width = 0.25)
plt.xticks(range(len(clubs)), clubs)
plt.xlabel("Club")
plt.ylabel("% of gender")
plt.legend(["male", "female"])
```

Out[154]: <matplotlib.legend.Legend at 0x12c808160>



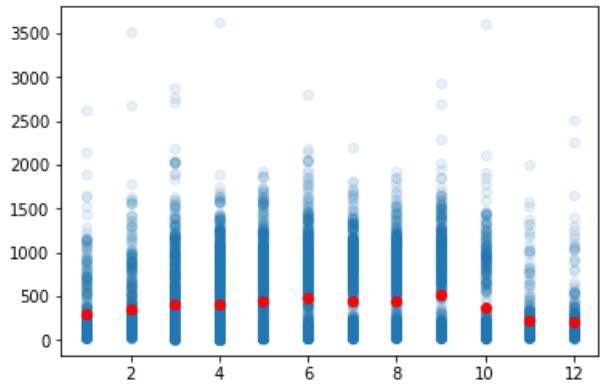
Analysis of Likes over Time

In [155]:
```
import csv
clubs = ["rudas", "oxfordsc_sd", "wetrepublic", "templeSF", "grandBoston", "grandSF", "prsymCH", "
data = pd.DataFrame()
for club in clubs:
    df = pd.read_csv(str("data_"+ club + "_posts.csv"), quoting=csv.QUOTE_NONE)
    data = data.append(df)
from datetime import datetime
dates = [datetime.strptime(date, '%Y-%m-%d %H:%M:%S') for date in data["DATE"]]
data["DATE"] = dates
```

In [156]:
```
# Distribution of likes for each post grouped by month. Red dot is the mean.
dates_month = [date.month for date in dates]
data["MONTH"] = dates_month
plt.scatter(data["MONTH"], data["LIKES"], alpha = 0.1)
for mean, index in zip(data.groupby("MONTH").mean()["LIKES"], range(1,13)):
    plt.plot(index, mean, 'ro')
print("Means:")
data.groupby("MONTH").mean()["LIKES"]
```

```
Means:
```

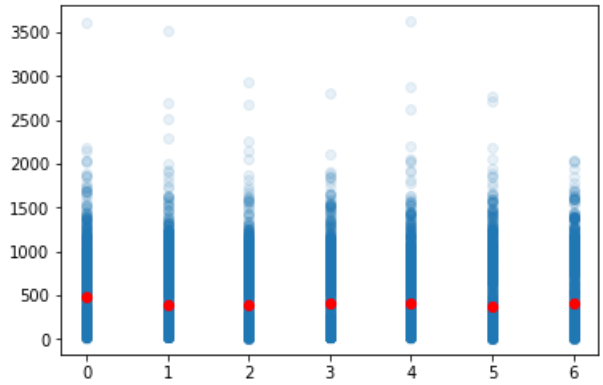
Out[156]:
```
MONTH
1     294.300971
2     357.642697
3     398.808974
4     415.246528
5     435.206385
6     477.146718
7     440.808696
8     449.480048
9     522.148670
10    379.797386
11    217.103448
12    213.541333
Name: LIKES, dtype: float64
```



In [157]:
```
# Distribution of likes for each post grouped by weekday (GMT). Red dot is the mean.
# where Monday is 0 and Sunday is 6
dates_day = [date.weekday() for date in dates]
data["day"] = dates_day
plt.scatter(data["day"], data["LIKES"], alpha=0.1)
for mean, index in zip(data.groupby("day").mean()["LIKES"], range(7)):
    plt.plot(index, mean, 'ro')
print("Means:")
data.groupby("day").mean()["LIKES"]
```

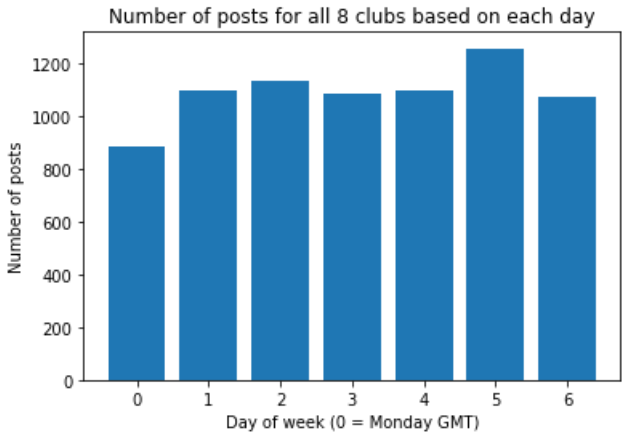
```
Means:
```

Out[157]:
```
day
0    486.562570
1    394.394353
2    382.904678
3    401.488029
4    404.617273
5    379.769658
6    411.043762
Name: LIKES, dtype: float64
```



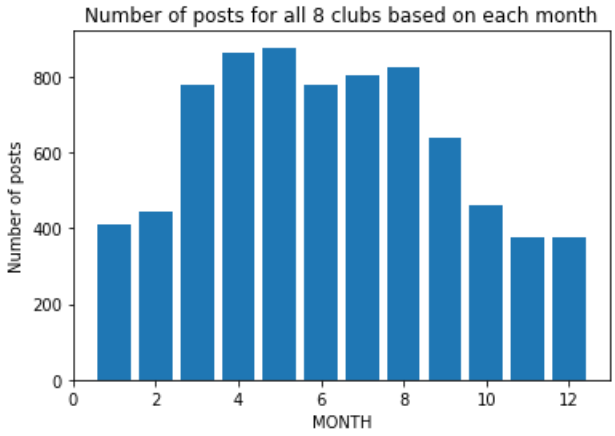
In [158]:
```
# Breakdown of posts by day
plt.bar(range(7), height= list(data["day"].value_counts().sort_index()))
plt.xlabel("Day of week (0 = Monday GMT)")
plt.ylabel("Number of posts")
plt.title("Number of posts for all 8 clubs based on each day")
```

Out[158]:
```
Text(0.5,1,'Number of posts for all 8 clubs based on each day')
```



In [160]:
```
# Breakdown of posts by day
plt.bar(range(1,13), height= list(data["MONTH"].value_counts().sort_index()))
plt.xlabel("MONTH")
plt.ylabel("Number of posts")
plt.title("Number of posts for all 8 clubs based on each month")
```

Out[160]:
```
Text(0.5,1,'Number of posts for all 8 clubs based on each month')
```



In [ ]: